ASPECTS OF THE MOLECULAR BIOLOGY

OF POTATO VIRUS Y

A thesis submitted in partial fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Molecular Biology

in the

University of Canterbury

by

Joanne M. Hay

University of Canterbury (1989)

Christchurch, New Zealand.

TABLE OF CONTENTS

ABBREVIATIONS

AI	amorphous inclusion
AlMV	alfalfa mosaic virus
APS	ammonium persulphate
ATP	adenosine triphosphate
A.tumefaciens	*Agrobacterium tumefaciens*
BAP	6-benzylaminopurine
BMV	bromes mosaic virus
bp	base pairs
BPB	bromophenol blue
BSA	bovine serum albumin
BYMV	bean yellow mosaic virus
CaMV	cauliflower mosaic virus
cDNA	complementary deoxyribonucleic acid
CIP	calf intestinal alkaline phosphatase
CI	cytoplasmic inclusion
CMV	cucumber mosaic virus
CPMV	cow pea mosaic virus
CP	coat/capsid protein
dd/dATP	dideoxy/deoxyadenosine triphosphate
dd/dCTP	dideoxy/deoxycytidine triphosphate
dd/dGTP	dideoxy/deoxyguanosine triphosphate
dd/dTTP	dideoxy/deoxythymidine triphosphate
DEPC	diethylpyrocarbonate
DIECA	diethyldithiocarbamic acid
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DTE	dithioerythritol
E.coli	*Escherichia coli*
EDTA	ethylenediaminetetra-acetic acid
ELISA	enzyme-linked immunosorbant assay
GA_3	gibberellic acid
HC	helper component
IPTG	isopropylthiogalactoside
JGMV	Johnson grass mosaic virus
kb	kilobases

M-MLV	murine myeloid leukemia virus
NAA	naphthaleneacetic acid
NI	nuclear inclusion
NOS	nopaline synthase
NPT II	neomycin phosphotransferase
OCS	octopine synthase
PEG	polyethylene glycol
PEBV	pea early browning virus
PeMV	pepper mottle virus
poly (A)	polyadenylate
PPV	plum pox virus
PRV	papaya ringspot virus
PWV	passionfruit woodiness virus
PTA	phosphotungstic acid
PVX	potato virus X
PVY	potato virus Y
RF	replicative form
RNA	ribonucleic acid
RUBISCO	ribulose bisphosphate carboxy oxygenase
SCMV	sugarcane mosaic virus
SDDC	diethyldithiocarbamic acid
SDS	sodium dodecyl sulphate
SMV	soybean mosaic virus
TCA	trichloroacetic acid
T-DNA	transfer deoxyribonucleic acid
TEM	transmission electron microscopy
Temed	N,N,N',N'-tetramethylethylenediamine
TEV	tobacco etch virus
TMV	tobacco mosaic virus
TRV	tobacco rattle virus
TVMV	tobacco vein mottle virus
VPg	viral genome linked protein
Xgal	5-bromo-4-chloro-3-indoyl-b-D-galactoside

ABSTRACT

Aspects from the coat protein from Potato Virus Y (PVY) were investigated. Polyadenylated, full-length RNA was isolated from purified preparations of two strains of the virus, PVY^N and PVY^C. The RNA was used as a template to produce double stranded complementary DNA (cDNA) which was susequently cloned into the plasmid vector pUC19. Of the resultant PVY^C-derived clones, two recombinant plasmids (pVYC5 and pVYC11) were analysed by DNA hybridisation and DNA sequencing. PVYC11 contained a viral cDNA insert, while pVYC5 did not.

A recombinant clone containing PVY^N cDNA sequences (pVYN27) was characterised by DNA sequencing. The 3'-terminal 1134 nucleotide sequence coded for the coat protein gene and contained one large open reading frame capable of encoding a protein of 264 amino acid residues with a combined molecular weight of 29 631. The ten amino-terminal amino acids of the protein were confirmed by amino acid sequencing.

Transcriptional fusions encoding the PVY^N coat protein gene and either the CaMV 35S promoter or the mannopine synthase promoter were inserted into an *Agrobacterium* binary vector encoding the NPT II gene. These were mobilised into two species of *Agrobacterium* (*A. tumefaciens*, LBA4404 and *A. rhizogenes*, A4T) and transformed into the genomes of *Nicotiana plumbaginifolia* and *Solanum tuberosum*. The transgenicity of regenerated tobacco plants was confirmed by 1. the presence of the chimaeric PVY^N coat protein gene as demonstrated by DNA hybridisation, 2. expression of the NPT II gene was demonstrated by the regeneration of plants on media containing kanamycin at normally inhibitory concentrations, and 3. demonstration that progeny from transformed plants inherited the NPT II gene in a Mendelian manner. No accumulation of the coat protein by transformants was detected by either Western blots or protein slot blots.

CHAPTER ONE

AN INTRODUCTION TO POTYVIRUSES AND A THESIS OVERVIEW

1.1. POTYVIRUSES

The potato Y viruses, or potyviruses, are the largest known group of plant viral pathogens (Matthews, 1981; Hollings and Brunt, 1981; Edwardson, 1974). They infect a wide variety of host plants including both monocotyledonous and dicotyledonous species. Most potyviruses have a narrow host range and this host specialisation minimises interspecific competition among members (Shukla and Ward, 1988a). Dependent on the definition of a distinct potyvirus, there are estimated to be more than 100 different members of this group (Edwardson, 1974; Francki *et al.*, 1985; de Bokx and Huttinga, 1981). Economic losses attributed to potyviral infection can be significant (de Bokx and Huttinga, 1981).

All potyviruses share a number of common features (Dougherty and Carrington, 1988). The virus particles are flexous rods approximately 700 to 900 nm long, and 12 to 15 nm in diameter. Potyviral genomes are monopartite and consist of a single stranded, positive sense infectious RNA of approximately 10,000 nucleotides, which constitutes 5% of the virion weight (Hill and Benner, 1976; Hinostroza-Orihuela, 1975; Allison *et al.*, 1986; Brakke and van Pelt, 1970; Domier *et al.*, 1986). The RNA genomes are 3'-polyadenylated and have a 5'-terminal genome-linked protein (VPg) (Hari, 1979, 1981). All potyviral genomes contain a protein which aggregates in the cytoplasm during infection to form characteristic 'pin-wheel' or 'scroll-shaped' inclusion bodies (Edwardson, 1974; Christie and Edwardson, 1977). Potyviruses can be mechanically transmitted, but most are transmitted by aphids in a non-persistent, non-circulative manner; transmission in seed and by other insect vectors has been reported (de Bokx and Huttinga, 1981). The closest relatives of the potyvirus group are the plant comoviruses and the animal picornaviruses, each of which possess RNA genomes that are linked to VPg molecules and encode large polyproteins (Carrington *et al.*, 1989).

Six virus-encoded proteins have been found associated with potyvirus infections: coat protein (CP), helper component (HC), VPg, cytoplasmic inclusions (CI) and two nuclear inclusion proteins (NIa and NIb). Potential cistrons capable of encoding two additional polypeptides (approximately 30 and 50 kd) are also present in the potyviral genome, but these proteins have not been identified *in vivo* (Hellmann *et al.*, 1986; Domier *et al.*, 1986; Allison *et al.*, 1986). The cistrons for these eight proteins have been mapped to the potyviral genome. Fig. 1.1 presents the current assessment of the genetic organisation of two potyviruses, tobacco etch virus (TEV) and tobaaco vein mottling virus (TVMV), based on a compilation of cell-free expression studies,

nucleotide sequence data and biochemical analyses of gene products (Dougherty and Carrington, 1988).

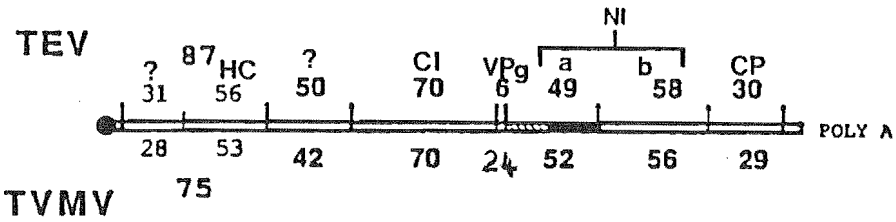


Figure 1.1. Proposed genetic maps for TEV and TVMV (after Dougherty and Carrington, 1988). The molecular weights (x 10^3) of the putative cleavage products are present for each virus. The following cistrons are identified: coat protein (CP), nuclear inclusion proteins (NIa & NIb), cytoplasmic inclusion protein (CI), helper component (HC) and VPg.

Gene products are initially expressed as a single, large polyprotein precursor that undergoes post-translational proteolytic processing, analogous to picornaviruses (Dougherty, 1983). The polyproteins which are produced have a molecular weight of approximately 340 kd (Allison *et al.*, 1986; Domier *et al.*, 1986; Hellmann *et al.*, 1983, 1986; Vance and Beachy, 1984; Yeh and Gonsalves, 1985). Intermediate cleavage products and individual mature viral proteins arise from the rapid proteolytic processing of the polyprotein.

The major structural potyviral protein is the coat protein and it has been mapped proximal to the 3'-terminus (Allison *et al.*, 1985b, 1986; Domier *et al.*, 1986; Dougherty and Hiebert, 1980c; Dougherty *et al.*, 1985, Ravelonandro *et al.*, 1988; Hay *et al*, 1989). It is the major component of the virion and envelopes the genomic material with approximately 2000 monomeric units (Hollings and Brunt, 1981). Each potyvirus contains a single type of capsid protein monomer, ranging in size from 30 to 45 kd (Allison *et al.*, 1985a). Heterogeneity in the apparent capsid protein size is a common feature of potyviruses owing to degradation during purification and/or storage (Shukla and Ward, 1988a). The amino acid sequences of a number of different potyviruses have been published (Shukla *et al.*, 1986, 1987, 1988b; Ravelonandro *et al.*, 1988; Dougherty *et al.*, 1985; Allison *et al.*, 1986; Domier *et al.*, 1986; Hay *et al.*, 1989). Analyses of assembled coat protein molecules by partial proteolytic digestion, biochemical analyses and immunological studies have revealed a number of similar structural features (Allison *et al.*, 1985a; Dougherty *et al.*, 1985; Shukla *et al.*, 1987, 1988c). Assembled potyviral coat proteins appear to have a structural organisation similar to potex- and tobamoviruses. The 3' position of the coat protein gene is unusual for viruses with a polyprotein mode of expression. In the picornaviruses and cowpea mosaic systems, capsid protein genes are located near the 5'-terminus, or are found

on a separate RNA species (Dougherty and Carrington, 1988). In TEV, there is evidence that the coat protein participates in the aphid transmission phenotype (Pirone and Thornbury, 1983).

A genome-linked VPg protein is found covalently attached to the 5'-terminal nucleotide of potyviral RNA from TEV (Hari, 1981) and TVMV (Shahabuddin *et al.*, 1988) virus preparations. The TVMV VPg is unusually large, 24 kd; for TEV the protein is 6 kd. A role for the VPg in replication can be inferred from studies of other viruses with similar gene organisation and expression (Dougherty and Carrington, 1988). The protein could conceivably be exposed externally on the virion and may have a potential role in aphid transmission (Dougherty and Carrington, 1988).

All potyviruses code for a cytoplasmic pin-wheel inclusion protein with similar molecular weights (65 to 75 kd) (Dougherty and Hiebert, 1980b; Edwardson, 1974). These aggregates consist of protein monomers and their morphology is virus-specific. No particular function has been conclusively assigned to this protein although it has been hypothesised to have a role in replication. The aggregates have also been observed to associate with the plasmodesmata (Edwardson, 1974) which implicates a role for this protein in the cell-to-cell movement of the virus. The CI protein may have multi-functional domains and play an integral role in both of these viral processes (Dougherty and Carrington, 1988).

A cytoplasmic amorphous inclusion body (AI) has been detected in a number of, but not all, potyvirus-infected plants. When formed, the polypeptide is large in size (51 to 56 kd) and is quite stable (de Mejia *et al.*, 1985a). Biochemical and immunological evidence exist that equate the AI and the helper component (de Mejia *et al.*, 1985b). The HC is a virus-encoded protein necessary for insect transmission (Berger and Pirone, 1986; Hiebert *et al.*, 1984; Pirone and Thornbury, 1983; Thornbury *et al.*, 1985; Thornbury and Pirone, 1983). However, purified AI bodies do not have HC activity suggesting that the inclusion-bound form of the protein is inactive. Alternatively, the HC activity may be associated with a processed or modified form of the inclusion protein (Dougherty and Carrington, 1988).

All potyviruses code for two proteins which aggregate in equimolar amounts (Dougherty and Hiebert, 1980b; Knuhtsen *et al.*, 1974). Together they comprise the nuclear inclusion body, although only a limited number of potyvirus infections cause stable NI aggregates to form (Edwardson, 1974). NI bodies vary in shape and size depending on the specific virus isolate. The two NI protein genes have been mapped adjacent to one another (Allison *et al.*, 1986; Hellmann *et al.*, 1986), and are approximately 58 and 49 kd, respectively. The 49 kd protein is a virus-encoded proteinase. It releases itself by an autocatalytic mechanism and functions in *trans* at three additional cleavage sites within the carboxy terminus of the TEV polyprotein (Carrington and Dougherty, 1987; Carrington *et al.*, 1988). On the basis of sequence homology studies, the 58 kd NI protein may be an RNA-dependent RNA polymerase (viral replicase) (Allison *et al.*, 1986; Domier *et al.*, 1987). The 3' positioning of the replicase to the VPg-proteinase cluster is consistent with observations made in other viral systems (Dougherty and Carrington, 1988).

The major cell-free translation products of TEV and TVMV RNA are proteins of 87 kd and 75 kd, respectively, and have been predicted to be encoded by sequences proximal to the 5' terminus of the open reading frame (Dougherty and Hiebert, 1980c; Hellmann *et al.*, 1980). These are thought to be precursors which are processed to produce proteins approximately 30 kd and 54 kd (Hellmann *et al.*, 1983, 1985; Hiebert *et al.*, 1984; Thornbury *et al.*, 1985). The smaller protein has been implicated in the cell-to-cell movement of the virus (Domier *et al.*, 1987). The remaining 54 kd portion has tentatively been identified as the aphid transmission HC and AI protein (Hellmann *et al.*, 1985; Hiebert *et al.*, 1984; Thornbury *et al.*, 1985). In TEV the proteinase responsible for the cleavage of this 5' precursor from the polyprotein has been identified and localised to the carboxy terminal half of the 56 kd HC (Carrington *et al.*, 1989). The proteinase appears to be released by an autocatalytic mechanism and cleavage *in vitro* occurs at a dipeptide which differs from those recognised by the 49 kd proteinase. Carrington *et al.* suggest the HC and the proteinase have separate domains within the 56 kd protein. Neither the site of cleavage between the 31 and 56 kd proteins nor the activity responsible for the cleavage in TEV have been identified. R. E. Rhoads (pers. comm.) observed that for TVMV two separate proteolytic activities are responsible for the maturation of the HC. One is encoded by the 5'-terminal gene (34 kd cistron) and mediates the HC/34 cleavage, and one is encoded by the HC cistron and mediates the HC/42 cleavage.

The gene located 3' to the sequence encoding the TEV 87 kd or the TVMV 75 kd 'precursor', is predicted to code for a protein with a molecular weight of approximately 50 kd (Domier *et al.*, 1987; Dougherty and Hiebert, 1980c). No specific function has been associated with this protein (Dougherty and Carrington, 1988).

There is no information available for the replicative cycle of potyviral RNA or on the manner of encapsidation.

1.2. THESIS OVERVIEW

Potato virus Y (PVY) is a member of the potyvirus group (Matthews, 1981) and causes economically important diseases world wide in potato, tomato, pepper and tobacco crops. Because of its prevalence in potato and tobacco crops in New Zealand and because few studies have been made of this 'type-member' for the potyvirus group, PVY was selected as the subject for this study. The initial aim of this study was to clone and sequence the PVY coat protein gene. Subsequently, the coat protein gene was to be expressed in host plants *Solanum tuberosum* and *Nicotiana plumbaginifolia*, with a view toward generating PVY-resistant plants.

Chapter one outlines some general properties of the potyvirus group and the current models for potyviral genome organisation and expression.

Chapter two reports the purification of two strains of PVY, a local isolate of the necrotic strain PVY^N, and an isolate of the PVY^C strain. DNA complementary (cDNA) to the genomic RNA from PVY^C was synthesised and cloned, and recombinant colonies were screened.

Chapter three reports the sequencing and subsequent analysis of the 3'-terminal 1134 nucleotides of the local isolate of PVY^N. A predicted amino acid sequence for the coat protein is inferred from the nucleotide sequence.

Chapter four reports the construction of chimaeric coat protein genes and their insertion into host plants *Nicotiana tabacum* and *Solanum tuberosum*, via the *Agrobacterium* binary vector for plant transformation. The presence of the PVY coat protein gene was examined by Southern blot analysis and DNA slot blot hybridisation. Expression of the gene *in vivo* was analysed by Northern and Western blots. The question of coat protein-induced 'genetically engineered cross protection' was addressed.

Chapter five summarizes the findings made by this study.

Each of chapters 2, 3 and 4 have separate introductions and conclusions as the area of study described in each, although complementary, is separate and complete.

CHAPTER TWO

VIRUS PURIFICATION AND CLONING OF THE COAT PROTEIN GENE

2.1. INTRODUCTION

2.1.1. SYMPTOMOLOGY

Potato Virus Y (PVY) is the type member of the potyvirus group (Matthews, 1981), and on the basis of symptom development on indicator plants comprises of three main groups of strains: PVY^O (common strain), PVY^N (tobacco veinal strain) and PVY^C (stipple streak strain) (de Bokx and Huttinga, 1981). Two of these, PVY^C and PVY^O, have long been identified in New Zealand, but PVY^N was isolated from potato only three years ago (Fletcher, 1986).

A PVY^N strain was first isolated in 1935 from tobacco growing near potato plants, in the United States of America (Smith and Dennis, 1940). Subsequently strains have been found all around the world, primarily in tobacco and potato crops. Although isolates from different countries were not found to be identical and often differed markedly in their physical characteristics, they were clearly related (Klinkowski and Schmelzer, 1960).

Differences in symptom expression in hosts depends on the strain of virus, cultivar of host plant and environmental conditions. Generally, primary infection by PVY^O strains induces severe systemic crinkle symptoms, leaf drop and premature death in potato cultivars; and yellowing, systemic mottling and leaf drop in tobacco. Potato cultivars susceptible to PVY^C strains show systemic mosaic, necrosis and stipple streak symptoms on the petioles and stems during primary infection, and may develop necrosis on the tubers during infection in the second year (secondary infection). Symptoms similar to those induced by PVY^O are observed for tobacco hosts. PVY^N strains cause necrotic rings and spots on the leaves of potato cultivars. The primary symptoms are usually mild, while those developing in the second year are more obvious. These extend from mosaic and streak on the leaves, to leaf drop and may lead to premature death of the plant. Severe systemic necrotic lesions, often following the main veins, are observed on the leaves of infected tobacco. Primary veins become brown and leaves collapse against the stem. The stems also show necrosis, especially near the base (Klinkowski and Schmelzer, 1960).

The PVY^N strains are highly aphid transmissable and all strains can be transmitted mechanically.

The economic significance of infection by PVY strains is severe as they cause devastating diseases in agronomically important crops. Klinkowski and Schmelzer (1960) state that 30 varieties of potato infected with PVY^C or PVY^O strains had the number of tubers produced reduced by an average of 39%, while yields were reduced 50.5%. Infection with necrotic strains

reduced tuber size by 12-14% on average, and yield by 14-20%. De Bokx and Huttinga (1981) quoted decreases of potato yields in the order of 10-80%, depending on host and viral strains. They also observed that the virus was transmitted to plant tubers during secondary infection. Infection of tobacco by PVY decreased yield by approximately 30% (Sievert, 1978), with PVY^N strains being particularly destructive and often causing complete crop failure. PVY also infects pepper and tomato cultivars which may result in heavy crop losses. The destructive effect of PVY is compounded when found in conjunction with other viruses, eg. potato leafroll virus and potato virus X (Edwardson, 1974). Hence, infection by PVY potentially incurs enormous annual financial losses to the agricultural sector in many countries throughout the world.

Nicotiana tabacum cvs. Samsun and White Burley are good indicator plants for differentiating strains of PVY; PVY^C infection results in vein clearing and slight epinasty of the leaves, and PVY^N produces necrotic lesions on the leaf and stem. The ease and rapidity of infection, and clarity of symptoms, contributed to *Nicotiana tabacum* varieties being good sources of virus. These varities were used as laboratory host plants for purification and subsequent experimentation.

2.1.2. CLONING

Recombinant DNA methods allow the isolation and amplification of single genes from whole genomes. They also facilitate the modification of genes which can then be re-introduced into cells for expression of RNA or protein (Wu, 1987). Such methods are useful for studying plant viruses, in particular RNA viruses. A copy of DNA (cDNA) complementary to the RNA viral genome can be made using reverse transcriptase (Hull and Davies, 1983). The cDNA can then be inserted into a prokaryote for replication and amplification. This approach contributes to an understanding of the complex problems of RNA virus evolution, the induction of disease symptoms, the function of plant viral sequences, the nature of virulence, and has contributed to the development of methods for reducing infection of plants.

cDNA has been used for sequencing the whole genomes of tobacco etch virus (TEV) (Allison *et al.*, 1986) and tobacco vein mottling virus (TVMV) (Domier *et al.*, 1986) and has led to a study of the relatedness of different virus species using nucleotide and amino acid sequence comparisons. Studies on the genetic organisation of plant viruses have been facilitated by using cDNA clones to the complete or partial genome, eg. pepper mottle virus (PeMV) (Dougherty *et al.*, 1985) and cowpea mosaic virus (CMV) (Vos *et al.*, 1984).

By recombining segments of viral genomes as DNA, phenotypes can be mapped physically to specific regions. Two early examples of this application are to the RNA viruses QB coliphage (Taniguchi *et al.*, 1978) and poliovirus (Rancaniello and Baltimore, 1981).

Disease symptom induction has been observed by initiating infection from cDNA clones to complete virus genomes (Turpen, submit.) It is also possible to synthesize infectious transcripts

in vitro from full-length cDNA clones of a number of plant RNA viruses including CPMV, brome mosaic virus (BMV) (Ahlquist *et al.*, 1984) and tobacco mosaic virus (TMV) (Meshi *et al.*, 1986).

cDNA fragments have been used as probes in nucleic acid hybridisation for identifying plant viruses, eg. citrus tristeza virus (Rosner *et al.*, 1984) and bean yellow mosaic virus (BYMV) (Abu-Samah and Randles, 1981). Rosner *et al.* (1986) cloned fragments of PVY to use as a diagnostic probe, via molecular hybridisation, for PVY infection.

The production of virus resistant plants using genetic engineering technology may potentially contribute to the prevention of plant diseases. For example, cDNA encoding the coat protein of TMV and alfalfa mosaic virus (AMV) has been expressed in plants via the *Agrobacterium tumefaciens* Ti plasmid, and has been shown to induce virus resistance (Powell-Abel *et al.*, 1986; Nelson *et al.*, 1987; van Dun *et al.*, 1988).

The first reports of molecular cloning of cDNAs were by Rougeon and Mach (1976) and Efstratiadis *et al.*, (1976). Since then, the techniques have been increasingly refined to provide versatile tools for the molecular analysis of eukaryote and prokaryote genes, examples of which have been described. Gubler and Hoffman (1983) modified the method of Okayama and Berg (1982) and combined classical oligo(dT)-primed first strand synthesis with the novel Rnase H-DNA polymerase I second strand synthesis. The production of double stranded cDNA became a simple two-step process. This procedure is useful because it permits synthesis from mRNAs which are present in low concentrations. It also eliminates the S1-nuclease mediated cleavage of hairpin loops generated by previous synthesis procedures, which usually removed important 5'-terminal sequences. D'Alessio *et al.* (1987) further modified the protocol by developing a 'one tube' system for the synthesis of cDNA from RNA by reverse transcription. A single strand of DNA complementary to the RNA template is sythesised from a primer by an RNA dependent DNA polymerase (reverse transcriptase). The primer is provided by a poly(dT) oligonucleotide annealed to the RNA 3'-poly(A) tail. The cDNA is made double stranded with DNA polymerase I from *E. coli* and the primer required by this enzyme is provided by short fragments of RNA generated by RNase H digestion of the cDNA-RNA heteroduplex (Gubler and Hoffman, 1983). The strategy of D'Alessio *et al.* is outlined in Fig. 2.1. and was used in this study. The aim was to produce complementary DNA to two strains of PVY, PVY^N and PVY^C. In this study, synthetic *Xba*I linkers were attached to the double stranded cDNA and the cDNA was inserted into the plasmid cloning vector, pUC19.

The pUC series of plasmids are widely used as cloning vectors. They encode ampicillin resistance and a portion of the β-galactosidase gene from the *E. coli* lac operon, including the promoter, which complements the portion encoded in *E. coli* hosts such as the JM83 strain. The pUC plasmids also contain a polylinker within the β-galactosidase encoding region, they propagate in high copy number, are small and do not contain sequences which inhibit transformation into eukaryote cells (Gronenborn and Messing, 1978). The detection of clones with inserts is via a simple colour assay. In the presence of an inducer,

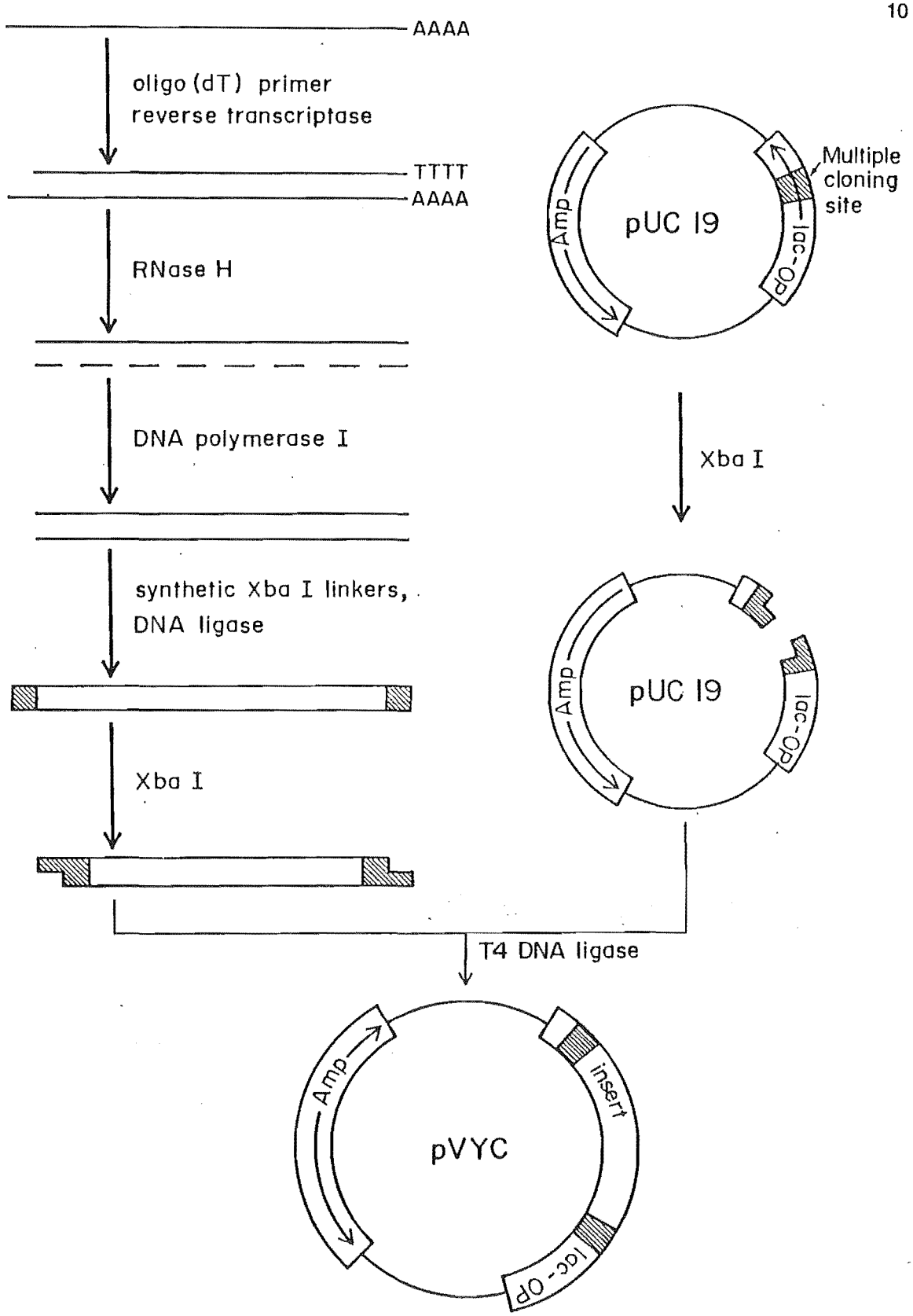


Figure 2.1. Strategy for cloning the PVY^N coat protein gene.

isopropylthiogalactoside (IPTG), the chromogenic β-galactosidase substrate, 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Xgal), is hydrolysed to bromochloroindole and confers a blue colour to the infected colony. If an insert is cloned into the polylinker, the α-peptide coding region is interrupted and no functional β-galactsidase can be produced, hence the recombinant colony is white.

In this study, the initial screening for recombinant PVY cDNA clones was by the expression of two recognisable phenotypes, ampicillin resistance and the loss of β-galactosidase activity. Further screening was by hybridisation with a nucleic acid probe known to encode PVY sequences.

2.2. MATERIALS AND METHODS

2.2.1. VIRUS ISOLATES AND PROPAGATION

Materials

PVY^N inoculum - isolated from *Solanum tuberosum* (Canterbury) and donated by J. Fletcher (Plant Diseases Division, Department of Industrial and Scientific Research, (PDD, DSIR), Lincoln, New Zealand), (Fletcher, 1986).

PVY^C inoculum - isolated from *Solanum tuberosum* cv. Lichte Rode Star and donated by J. A. de Bokx, (Research Institute for Plant Protection, Wageningen, The Netherlands).

Methods

Fresh or freeze-dried inoculum was homogenised in an oven-baked mortar and pestle in four to five volumes (w/v) of Yarwoods buffer (0.5% potassium phosphate, 0.5% bentonite; Yarwood 1968). Both strains of the virus were propagated in *Nicotiana tabacum* cv. White Burley by mechanical inoculation. Four to six week old plants were dusted with carborundum powder (400+ mesh) on two to three large leaves per plant and rubbed with the inoculum. Excess inoculum and carborundum were immediately rinsed off with distilled water to prevent dessication.

Plants were grown in insect-free growth rooms or glasshouses at 18^{O}-20^{O}C. Symptoms developed 10-21 days after inoculation, depending on the strain, and infection was confirmed using a commercial Enzyme Linked Immunosorbent Assay (ELISA) kit (Boehringer Mannheim). Samples were assayed according to the manufacturers instructions and the intensity of the colour reaction assessed visually. A value from - to + + + was assigned.

2.2.2. VIRUS PURIFICATION

Four protocols for the purification of PVY were used including those of Hamilton and Nichols (1978), R.Forster (pers. comm.) and Dougherty and Hiebert (1980a). The highest yield of intact virus particles were purified using a procedure published by Reddick and Barnett (1983) with modifications to the initial extraction buffer.

Fresh leaves showing either severe veinal necrosis (PVY^N infected) or mild mosaic mottling (PVY^C infected) were homogenised in 1.5 volumes (w/v) of phosphate buffer (0.5M potassium phosphate pH 7.0, 1.0M urea, 0.01M EDTA, 0.01M diethyldithiocarbamic acid (SDDC), 0.1% ascorbic acid) in a Waring blender at high speed for two to three minutes. Chloroform was added (0.8 volumes (v/v)) and the homogenate blended for a further one minute. The emulsion was filtered through cheesecloth and the supernatant centrifuged at 8000 rpm for fifteen minutes

in a Sorvall GS-3 rotor. The resulting pellets were re-extracted in 0.75 volumes of phosphate buffer and 0.4 volumes of chloroform and centrifuged as before.

The virus was concentrated from the combined supernatants by adding 4% polyethylene glycol (PEG) (w/v) and 0.25M NaCl and stirring for one hour at 4°C. The PEG precipitate was collected by centrifugation at 8000 rpm for 15 minutes and resuspended in 50.0 mls of phosphate/urea buffer (0.5M potassium phosphate pH 7.0, 1M urea) containing 1% Triton X-100. The suspension was stirred for two hours at 4°C and then centrifuged at 8000 rpm for 15 minutes in a Sorvall SS-34 rotor. The supernatant was subjected to a second PEG precipitation by stirring with 4% PEG (w/v) and 0.25M NaCl for one hour at 4°C, and the crude virus pellet was collected by centrifugation at 8000 rpm for 15 minutes.

Equilibrium centrifugation in caesium sulphate (Cs_2SO_4) followed immediately: the virus pellet was resuspended in 7.5 ml of phosphate/urea buffer containing 1.5g of Cs_2SO_4, and layered onto two 0.8 ml cushions of 53% (w/w) Cs_2SO_4 in 20mM Tris-HCl pH 7.5 in two Beckman SW55 ultraclear centrifuge tubes. The tubes were topped up with phosphate/urea buffer and centrifuged at 30,000 rpm at 5°C for 16 hours in a Beckman SW55 rotor. The opalescent virus band was removed from the Cs_2SO_4 using a peristaltic pump, diluted five times in 20 mM Tris-HCl pH 7.0, and dialysed against one litre of 20mM Tris-HCl pH 7.0 for four hours with a buffer change after one hour.

The resulting suspension was cleared by centrifugation at 10,000 rpm for 10 minutes in a Sorvall SS34 rotor and the virus collected from the supernatant by high speed centrifugation at 50,000 rpm for 1.5 hours in a Beckman SW55 rotor. The glassy pellet was resuspended in 250 μl of 20mM Tris-HCl pH 7.0.

The presence of virus was ascertained by transmission electron microscopy (TEM) using 2% phosphotungstic acid (PTA) pH 7.0 as a negative stain. PVY concentrations were determined spectrophotometrically using a coefficient of 2.3 mg ml^{-1} at 260nm or by the dye-binding method of Bradford (1976). A spectrophotometric profile between 220nm and 320nm was produced.

Whole virus was stored in aliquots at -20°C.

2.2.3. RNA ISOLATION

Freshly purified virus was dissociated as described by Brakke and van Pelt (1970). Virus particles were dissolved in a small volume of disruption buffer (100mM Tris-HCl pH 9.0, 1mM EDTA, 1% sodium dodecylsulphate (SDS), 0.1 mg ml^{-1} bentonite) containing 10 μg ml^{-1} proteinase K (pre-incubated at 37°C for 10 minutes). After gentle mixing, the mixture was incubated at room temperature for 15 to 30 minutes.

Sucrose Gradient Centrifugation (Brakke and van Pelt, 1970)

Two linear 7.5%-30% sucrose gradient columns were prepared in 50mM Tris-HCl pH 9.0, in SW28 polyallomer centrifuge tubes using a gradient former. One ml of disrupted virus was layered onto

the top of one gradient, the tubes topped with paraffin oil to prevent collapse and centrifuged at 25,000 rpm for 13 hours at 14^{o}C.

The tubes were pierced using a Beckman fraction recovery system and one ml fractions were collected. The absorbance of each fraction was measured at 260nm and the peak fractions pooled and precipitated. The RNA was resuspended to 1 mg ml^{-1} in double glass distilled (dd)H_2O and stored at -80^{o}C in 10 μl aliquots.

The length and size distribution of the resulting RNA was determined by gel electrophoresis in 1% agarose gels containing 1x TBE (section 2.2.13.).

2.2.4. SYNTHESIS OF COMPLEMENTARY DNA

Double stranded complementary DNA (cDNA) was synthesised from 5 ug of PVY RNA using the single tube reaction described by d'Alessio *et al.* (1987).

First Strand Synthesis

A 50 μl reverse transcription reaction was prepared containing 5 μg of PVY RNA heat treated to 65^{o}C for 10 minutes and quick chilled, plus 50mM Tris-HCl pH 8.3, 75mM KCl, 10mM dithioerythritol (DTE), 3mM $MgCl_2$, 500 μM each of dATP, dGTP, dCTP and dTTP, 50 μg ml^{-1} oligo $(dT)_{12-18}$ primer and 10,000 units ml^{-1} M-MLV reverse transcriptase.

After the enzyme was added, the reaction mixture was kept on ice and a 10 μl aliquot transferred to a separate tube containing 10 μCi of [α-^{35}S]dCTP (1000 Ci mmol^{-1}) or 7.6 μCi of [methyl-^{3}H]dTTP (47 Ci mmol^{-1}). A one μl sample of this labelled reaction was immediately removed, TCA precipitated (section 2.2.11) and counted to provide a 'time zero' ('t_0') product. Both tubes were incubated at 37^{o}C for one hour; the reaction was stopped by placing the tubes on ice. A one μl sample was removed from the labelled reaction, TCA precipitated and counted to determine the yield of first strand product at 't_{60}'.

Second Strand Synthesis

The tube containing 40 μl of unlabelled first strand reaction was diluted to a final volume of 320 μl containing 25mM Tris-HCl pH 8.3, 100mM KCl, 5mM $MgCl_2$, 5mM DTE, 250 μM each of dATP, dGTP, dCTP, dTTP, 500 cpm/pmole [α-^{35}S]dCTP (or 250 cpm/pmole [methyl-^{3}H]dTTP), 250 units ml^{-1} *E. coli* DNA polymerase I and 8.5 units ml^{-1} *E. coli* RNase H. The mixture was incubated for two hours at 16^{o}C and synthesis was terminated by the addition of 10 μl of 0.5M EDTA.

Five μl aliquots were removed at 't_0' and 't_{120}', TCA precipitated and counted to determine the yield of second strand cDNA. The remainder of the reaction mixture was extracted with an equal volume of phenol/chloroform and concentrated by ethanol precipitation using ammonium acetate (section 2.2.9.). The pellet was resuspended in either 25 μl TE (10mM Tris-HCl pH 8.0,

1mM EDTA) or 5 μl of 1x T4 polynucleotide kinase buffer (section 2.2.5.) in preparation for ligation of linkers, and stored at 4^{O}C.

Portions of single and double stranded cDNAs were analysed by 1% agarose gel electrophoresis and the remainder was used for cloning.

2.2.5. CLONING OF COMPLEMENTARY DNA INTO A PLASMID VECTOR

Blunt End Ligation of Synthetic Linkers

Synthetic *Xba*I linkers were supplied in 5'-hydroxyl, single stranded form (Collaborative Research). They were re-annealed by heating to 65^{O}C for 10 minutes and allowed to cool slowly to room temperature. The double stranded linkers were phosphorylated at their 5'-hydroxyl ends as follows: one μg of double stranded *Xba*I linkers was incubated at 37^{O}C for one hour in 10 μl of kinase buffer (10mM DTE, 10mM Tris-HCl pH 7.5, 1mM ATP, 10mM $MgCl_2$, 200 ug ml^{-1} acetylated bovine serum albumin (BSA)) and two units of T4 polynucleotide kinase (Maniatis *et al.*, 1982).

This reaction was added to 10 μg of cDNA in 5 μl of 1x kinase buffer for blunt end ligation. After the addition of two units of T4 DNA ligase and 0.4 μl of 17.7mM ATP, the reaction was incubated at 22^{O}C for 6-10 hours. The linker ligation reaction was stopped by the addition of 2 μl of 0.5M EDTA pH 8.0, and extracted once with phenol:chloroform:isoamyl alcohol, (25:24:1). The organic phase was back-extracted once with 20 μl of TE and the combined aqueous phases precipitated with ethanol. The dried pellet was resuspended in 100 μl of buffer (50mM Tris-HCl pH 8.0, 10mM $MgCl_2$, 50mM NaCl), 100 units of *Xba*I restriction endonuclease were added and incubated at 37^{O}C for four hours. The reaction was stopped by adding 2 μl of 0.5M EDTA, extracted once with phenol/chloroform and the excess linkers removed by gel filtration on a spun Sephadex G-50 column (section 2.2.10.). Five μl samples were taken from the cDNA after spinning through the column, TCA precipitated and counted to determine losses of cDNA.

The linkered cDNA eluate was ethanol precipitated and resuspended in 10 μl of TE pH 7.0 for ligation into vector. A one μl sample was counted to determine the final yield.

Preparation of Vector for Ligation

Ten μg of plasmid vector pUC19 were made linear in a 100 μl reaction by digestion with *Xba*I at its single recognition sequence in the multiple cloning site.

To minimise the number of non-recombinant molecules being produced during ligation, 5'-phosphate groups were removed from the linear plasmid by reaction with alkaline phosphatase (Maniatis *et al.*, 1982). Following digestion with *Xba*I, pUC19 DNA was dissolved in a 50 μl reaction containing 50 mM Tris-HCl pH 9.0, 1mM $MgCl_2$, 0.1mM $ZnCl_2$, 1mM spermidine and 0.1 unit of calf intestinal alkaline phosphatase (CIP). The reaction was incubated for one hour at 37^{O}C with a second aliquot of CIP being added after 30 minutes. The reaction was then diluted

to 100 μl in STE (10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA) and 0.5% SDS. The CIP was inactivated by heating the reaction mix to 65^{o}C for 15 minutes, extracted twice with phenol/chloroform and the DNA precipitated in ethanol. The plasmid DNA was resuspended in 10 μl TE pH 7.5 in preparation for ligation to the cDNA.

Ligation of Vector and cDNA

Vector pUC19 DNA and linker treated cDNA were incubated at 16^{o}C overnight in a 10 μl ligation reaction (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 10mM DTE, 1mM spermidine, 1mM ATP, 100 μg ml^{-1} BSA) with one Weiss unit of T4 DNA ligase.

A total of 120 ng of DNA, and a vector to insert ratio of 5:1 in the 10 μl reaction was used. A control reaction containing no cDNA was included to determine the efficiency of the dephosphorylation of pUC19.

Transformation

E. coli strain JM83 cells were made competent and transformed with the pUC19/cDNA ligation reaction using a modification of the method published by Hanahan (1983). Ten ml of SOB (2% (w/v) tryptone, 0.5% yeast extract, 10mM NaCl, 10 mM $MgSO_4$, 10mM $MgCl_2$, 2.5mM KCl) were inoculated with a single colony of JM83 and incubated overnight at 37^{o}C with agitation. One ml of this culture was used to inoculate 10-30 mls of SOB in a 250 ml flask and allowed to incubate at 37^{o}C with shaking until an absorbance of 0.45-0.55 at 550 nm (A_{550}) was reached. The cells were collected in a pre-chilled 50.0 ml polypropylene centrifuge tube, placed on ice for 10-15 minutes and pelleted at 2500 rpm for 12 minutes at 4^{o}C in a Sorvall SS34. The cells were resuspended in TFB buffer (10mM KMES pH 6.2, 100mM RbCl, 45mM $MnCl_2.4H_2O$, 10mM$CaCl_2.2H_2O$) at one third of the initial volume with gentle vortexing and placed on ice for 15 minutes. The cells were pelleted as before and resuspended in TFB buffer at 1/12.5 of the initial volume of cells. Fresh dimethyl sulphoxide (DMSO) was added to 3.5% and the cells left on ice for 5 minutes. DTE was added to 7.5%, gently swirled and the cells incubated on ice for a further 10 minutes. A second portion of DMSO was added and the cells iced for 5 minutes. 210 μl samples of competent cells were placed in pre-chilled eppendorf tubes and the 10 μl ligation mixtures added, gently mixed and incubated on ice for 30 minutes. The tubes were heat pulsed at 42^{o}C for 90 seconds and placed on ice for 1-2 minutes. 800 μl of SOC (SOB containing 20mM glucose) was added to each tube and allowed to shake slowly at 37^{o}C for one hour. The cells were pelleted in a microfuge for 5 minutes, resuspended in 200 μl of SOB and spread on LM plates (1% tryptone, 0.5% yeast extract, 10mM NaCl, 10mM $MgSO_4.7H_2O$, 1.5% (w/v) bactoagar) containing 50 μg ml^{-1} ampicillin and 0.1 mg ml^{-1} Xgal. The plates were incubated at 37^{o}C to establish colonies.

2.2.6. COLONY SCREENING

Colonies that were ampicillin resistant and lac$^-$ (unable to metabolise Xgal) were screened for insert size by making mini-preparations of the recombinant plasmid DNA using a rapid alkaline extraction procedure (Birnboin and Doly, 1979), followed by digestion with *Xba*I and analysis of insert size by electrophoresis through 1% agarose gels.

Colonies to be screened were dispersed in one ml of LB media (1% tryptone, 0.5% yeast extract, 1% NaCl) containing 50 μg ml^{-1} ampicillin, in eppendorf tubes and incubated overnight at 37^{o}C. The cells were pelleted in a microfuge, iced and resuspended in 100 μl of pre-lysis buffer (50mM glucose, 25mM Tris-HCl pH 7.6, 1mM EDTA) by vigorous vortexing. Cells were lysed by adding egg white lysozyme to 2.5 mg ml^{-1} and were incubated at room temperature for 15 minutes. 200 μl of freshly prepared lysis solution (200mM NaOH, 1% SDS) were added to the cells, mixed gently by inversion and allowed to incubate for 5 minutes at room temperature. 150 μl of acidifying buffer (3M NaOAc pH 4.8) were then added, mixed by inverting the tubes 30 times and incubated on ice for 30 minutes. Insoluble debris was pelleted by microfugation for 15 minutes and the supernatant decanted into a new tube. One ml of ethanol was added and the tubes centrifuged at room temperature for 3-4 minutes. After drying the tubes to remove residual ethanol, the pellets were resuspended in neutralising buffer (0.1M NaOAc, 50mM Tris-HCl pH 8.0) by vigorous vortexing and incubated on ice for 2-3 minutes. This was centrifuged to precipitate any remaining proteins and SDS, the supernatant retained and the nucleic acids precipitated in 700 μl of ethanol. The tubes were centrifuged for five minutes, the ethanol poured off and the pellet dried and resuspended in 50 μl TE pH 7.5 containing 10 μg ml-1 of RNase A.

Ten μl samples were loaded on a 1% agarose gel to visualise the recombinant plasmids. Twenty μl samples were digested with 10 units of *Xba*I restriction endonuclease for 30 minutes at 37^{o}C (section 2.2.5). *Xba*I/*BamH*I restriction endonuclease double digests were incubated in high salt conditions with 10 units each of *Xba*I and *BamH*I and digested for 30 minutes at 37^{o}C. Where *Xba*I/*Sca*I double digests were needed, samples were digested with 10 units of *Sca*I and 6mM Tris-HCl pH 7.5, 6mM $MgCl_2$, 150mM NaCl, 0.1 mg ml^{-1} BSA for 30 minutes at 37^{o}C, then adjusted to high salt buffer conditions. Ten units of *Xba*I were added and allowed to digest for a further 30 minutes. These samples were analysed on a 1% agarose gel for cDNA insert lengths.

2.2.7. DOT BLOT HYBRIDISATION

Plasmids were purified in large scale from selected colonies by the method of Coleman *et al.* (1978). Recombinant plasmids were bound to Genescreen Plus hybridisation membrane (Du Pont) and probed with a ^{32}P-labelled cDNA insert fragment isolated from pVYN27, known to encode internal sequences of the PVY^{N} 3'-terminal coat protein gene. This was to confirm whether the inserts were of viral origin.

Probe Preparation

An internal *Cla*I restriction endonuclease digested fragment from pVYN27 was purified on a 6% polyacrylamide gel as described (section 3.2.1.). Twenty-five ng of fragment were labelled with [α-^{32}P]dCTP using the Amersham Multiprime DNA labeling system as described in section 4.2.6.

DNA Slot Blot Hybridisation

Twelve pg of purified plasmid DNA and 5 μg of crudely prepared whole plant DNA from PVY-infected and uninfected *Nicotiana tabacum* was blotted onto the 'B' side of a piece of Genescreen Plus membrane in a 200 μl volume of 0.125N NaOH, 0.125x SSC (0.01875M NaCl, 0.001875M Na citrate) using a Bio-dot SF blotting apparatus (Biorad). Non-recombinant pUC19, pBR322, pGQ101, recombinant pVYN27 and unlabelled *Cla*I fragment isolated from pVYN27 were used as controls.

The membrane was neutralised, hybridised and washed as described in section 4.2.7. The membrane was sealed in a plastic bag and used to expose an X-ray film at -80^{o}C for three days as described by Maniatis *et al.* (1982).

2.2.8. SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS

One μg of PVY^{N} whole virus was mixed with an equal volume of Laemmli (1970) sample buffer (0.0625M Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 5% 2-β-mercaptoethanol, 0.001% bromophenol blue) and heated for three minutes in boiling water before loading onto the gel. Electrophoresis was performed on a 10% polyacrylamide separating gel (375mM Tris-HCl pH 8.8, 10% acrylamide:bis (29.2:0.8), 0.1% SDS, 0.1% ammonium persulphate, 15 μl N,N,N'-tetra methylethylene diamine (TEMED)) and run in electrophoresis buffer (0.025M Tris-HCl pH 8.3, 0.192M glycine, 0.1% SDS) according to the method of Laemmli. The Biorad markers used were low-range molecular weight protein markers and were phosphorylase b (97,400 d), bovine serum albumin (66,200 d), ovalbumin (42,699 d), bovine carbonic anhydrase (30,000 d), soybean trypsin inhibitor (21,500 d) and lysozyme (14,400 d). Following electrophoresis, the gel was stained with either a coomassie blue stain (0.125% Coomassie Blue R-250, 50% methanol, 10% acetic acid) for half an hour and destained in 5% methanol, 10% acetic acid, or with the more sensitive silver stain (Switzer *et al.*, 1979; Oakley *et al.*, 1980).

2.2.9. PHENOL/CHLOROFORM EXTRACTION AND ETHANOL PRECIPITATION (MANIATIS *et al.*, 1982)

Nucleic acid solutions were deproteinised by mixing with an equal volume of phenol saturated with 0.1M Tris-HCl pH 8.0. Phases were separated by spinning in a microfuge for five minutes at 13,000 rpm, and the upper aqueous phase collected. The organic phenol layer was back-extracted with an equal volume of TE pH 8.0 (or ddH_2O), the phases separated as before and the aqueous phases pooled. A second equal volume of phenol/chloroform (buffer saturated

phenol:chloroform:iso-amyl alcohol, 25:24:1) was added to further deproteinise, followed once by two volumes of chloroform (chloroform:iso-amyl alcohol, 24:1).

Nucleic acids were precipitated from the pooled aqueous solution by adding 0.1 volume of 3M NaOAc pH 5.0 (to 0.3M), and two volumes of cold, 100% ethanol (to 70%). The solutions were incubated either overnight at 4°C or for 15 minutes at -80°C, and the nucleic acids pelleted by centrifugation at 13,000 rpm for 10 minutes at 4°C. The supernatant was poured off and traces of ethanol removed by drying the pellet *in vacuo*. If the pellet showed white salt residues, the pellets were gently washed with a further aliquot of cold ethanol and pelleted as before.

For quantitative recovery of less than microgram quantities, DNA was precipitated by adding 0.5 volumes of 7.5M NH_4OAc (to 2.5M) and 2.5 volumes of 100% ethanol (to 70%) (Crouse and Amorese, 1987). The precipitate was allowed to form overnight at 4°C and collected by microfugation for 30 minutes. The pellet was dried and washed as above.

2.2.10. SPUN SEPHADEX G-50 COLUMN (MANIATIS *et al.*, 1982)

A spun Sephadex G-50 column was used for size fractionation of DNA samples. A one ml disposable plastic syringe plugged with a disc of porous polyethylene was filled with a slurry of G-50 beads (0.1g G-50 dissolved in STE). The column was spun in a glass conical centrifuge tube for four minutes at 3400 rpm in a bench top centrifuge to pack the column to a volume of 0.9 ml. The column was rinsed by adding 100 μl of STE and re-centrifuging; this was repeated twice more. The DNA was then added in a 100 μl volume, centrifuged as before and the effluent collected in a capless eppendorf tube in the bottom of the glass centrifuge tube.

2.2.11. TRICHLOROACETIC ACID (TCA) PRECIPITATION (MANIATIS *et al.*, 1982)

The yields of a number of reactions were calculated by measuring the radioactivity of small samples of the labelled DNA. Labelled DNA was separated from unincorporated labelled nucleotides by precipitation in excess cold 10% TCA and 0.1% Na pyrophosphate, and collected by filtration onto glass fibre filters (Whatman GF/C). The filters were rinsed with cold 5% TCA and dried at 80°C for 15-30 minutes. The dried discs were placed in plastic scintillation vials containing 5.0 mls of scintillation fluid (toluene/POPOP/PPO). Samples were counted in a Philips PW4540 scintillation counter on the appropriate channel and yields were calculated based on the known specific activities for each reaction.

2.2.12. PREPARATION OF BENTONITE

Bentonite is a macaloid known to adsorb RNase. It was prepared as follows (Maniatus *et al.*, 1982): 0.5 g of bentonite powder was suspended in 50.0 mls of sterile 50mM Tris-HCl pH 7.6 and heated to 100°C for five minutes with constant agitation. The solution was centrifuged at room temperature at 2500 xg for 15 minutes and the pellet resuspended in 40.0 mls of sterile 50mM Tris-HCl pH 7.6. The washing and centrifugation steps were repeated twice and the final

suspension was centrifuged at 3500 xg for 15 minutes. The final pellet was resuspended in 30.0 ml of sterile 50mM Tris-HCl pH 7.6, giving a final concentration of bentonite as 16 mg ml^{-1}. This can be stored indefinately at 4^{o}C.

2.2.13. GEL ELECTROPHORESIS

The lengths of various RNA and DNA species were estimated by gel electrophoresis. One percent agarose mini-gels made in 1x TBE (89nM Tris-borate, 89mM boric acid, 8mM EDTA) were used for analysis of viral RNA, single stranded and double stranded cDNA and for screening recombinant plasmids.

RNA gels were run and prepared in ribonuclease- (RNase) free conditions. Ten to twenty μl samples were mixed with 5x sample buffer (50% glycerol, 0.25% bromophenolblue in 5x TBE) and loaded onto each lane. Molecular weight markers were one μg of 1 kb ladder (BRL) and electrophoresis was run in 1x TBE at 5 V cm^{-1}. After the blue dye was seen to have migrated approximately one third to one half the way down the gel, it was removed and stained in a 0.5 μg ml^{-1} solution of ethidium bromide for 15 minutes. Gels were then examined on a short wavelength ultraviolet transilluminator and photographed with Polaroid 667 film for ten second exposures.

2.2.14. ENZYME TREATMENT

Various RNA and DNA preparations were examined for composition by treatment with either DNase or RNase A.

DNase Treatment

Twenty μl of nucleic acid was incubated with 2 μg of RNase-free DNase (BRL) and 50mM NaOAc pH 6.5, 10mM $MgCl_2$, 2mM $CaCl_2$, for one hour at 37^{o}C.

RNase A Treatment

A one mg ml^{-1} stock of RNase A (Sigma) was heated to 100^{o}C for 15 minutes and allowed to cool slowly to room temperature. Two μg were added to 20 μl of nucleic acid in 20mM Tris-HCl pH 7.5 and incubated for 30 minutes at room temperature.

2.2.15. STERILITY

All solutions, glassware and equipment used for working with RNA were ribonuclease free. Where possible, items and solutions were autoclaved for 15 minutes at 121^{o}C and 15 psi. Sterile non-autoclavable solutions and plasticware were treated with 0.1% diethylpyrocarbonate (DEPC) overnight at room temperature and then rinsed with sterile dH_2O. Glassware, eppendorf tubes and Gilson pipette tips were autoclaved and baked at 80^{o}C overnight. Disposable gloves were worn when conducting experiments involving RNA.

Bacteriological sterility was maintained for solutions and media by autoclaving or by filtration through sterile Millipore 0.22 μm filters. Standard aseptic techniques were used.

2.3. RESULTS

2.3.1. VIRUS PROPAGATION

Nicotiana tobacum plants inoculated with PVY^N developed typical severe veinal necrotic symptoms on leaves and necrotic lesions on the stem, accompanied by chlorotic yellowing of the leaves after 11-14 days. Plants inoculated with PVY^C developed mild mottling on the leaves and some veinal clearing around the primary and secondary leaf veins 14-21 days after inoculation. (Plates 2.1. and 2.2.).

Samples of leaves inoculated with PVY^N were analysed by ELISA and revealed an increase in viral titre in systemically infected leaves from day 1-10. No virus was detected in directly inoculated leaves until day 6. By day 10, the concentration of virus appeared similar in both directly inoculated and systemically infected leaves. (Table 2.1.).

Table 2.1. Colour development in ELISA tests on PVY^N systemically infected and directly inoculated leaves from *N. tabacum*.

	DAYS									
	1	2	3	4	5	6	7	8	9	10
Dir. inoculated	-	-	-	-	-	+	++	++	+++	+++
Systemic	+	+	+	+	++	++	++	+++	+++	+++

2.3.2. VIRUS PURIFICATION

Four potyviral purification protocols were tested for final PVY yield. All four procedures used an initial extraction in buffered solution containing a number of antioxidants and organic solvents, followed by a low speed centrifugation to separate plant debris and organic and aqueous phases, with the virus partitioning into the aqueous phase.

The methods published by Hamilton and Nichols (1978) and Dougherty and Hiebert (1980a) use low salt buffers for the initial extraction of virus from infected material: 0.1M Tris-HCl pH 9.0 and 20mM Hepes pH 7.0, respectively. Methods by Reddick and Barnett (1983) and R. Forster (pers. comm.) use high salt buffers: 0.5M KPO_4 pH 7.0, 1M urea and 0.5M KPO_4, respectively. Samples taken from the initial homogenate before and after low speed centrifugation and tested by ELISA, showed higher yields of virus in the supernatant when a high ionic strength buffer was used. All procedures showed a significant loss of virus in the pellet after centrifugation.

Plate 2.1. *Nicotiana tabacum* cv. White Burley, 21 days after inoculation with PVY^{C}. Veinal clearing and mild chlorotic mottling can be observed. **(A)** Whole infected plant. **(B)** Comparison of uninfected (un) and infected (in) leaves.

Plate 2.2. *Nicotiana tabacum* cv. White Burley, 14 days after inoculation with PVY^N. Veinal and stem necrosis, and general chlorosis of the leaf can be observed. **(A)** Whole infected plant. **(B)** Infected leaf (after A. Fellowes, MSc. thesis, 1988).

The method giving the best overall recovery was that described by Reddick and Barnett (1983) and this was used for all PVY^N and PVY^C purifications. Virus was purified from the aqueous phase of the initial homogenate by a series of PEG precipitations; viral pellets were resuspended in buffers containing Triton X-100; the virus was finally purified by isopycnic centrifugation in a Cs_2SO_4 gradient. These steps in particular, facilitated the purification of a virus prone to aggregation. The virus formed an opalescent band in the Cs_2SO_4 gradient (Plate 2.3.) and after dialysis and high speed centrifugation, the final viral pellets were clear.

Whole virus was electrophoresed on a 10% SDS polyacrylamide gel and was stained with coomassie blue or silver stain (Plate 2.4.). The gels showed two protein bands at approximately 34 and 31 kilodaltons (kd) by comparison with protein markers. The larger was in good agreement with published PVY coat protein molecular weights (de Bokx and Huttinger, 1981, Shukla *et al.*, 1986). The preparations were shown to be free of contamination by other proteins. This was supported by an observed A_{260}:A_{280} ratio of 1.3, which is in agreement with the published value of 1.22 for pure PVY (de Bokx and Huttinga, 1981). Examination of a PTA stained preparation using a transmission electron microscope showed the virus particles to be slighty aggregated and to have a modal length of 750 nm (Plate 2.5.).

Yields were estimated as 1-5 mg kg^{-1} of infected plant material, based on 1 mg ml^{-1} of purified PVY having an absorbance at 260 nm (A_{260}) of 2.3 (de Bokx and Huttinga, 1981). An example of an absorbance profile for PVY^C is shown in Fig. 2.2.

2.3.3. RNA ISOLATION

A minimum of 5 mg of purified virus was required before detectable RNA could be isolated. In most RNA preparations, the A_{260} readings of fractions collected from the sucrose density gradients revealed a major absorbance peak, as shown in Fig. 2.3. The A_{260}/A_{280} of the pooled fractions of PVY^C RNA, from the major peak in Fig. 2.3., was 1.7. The integrity of the isolated viral RNA is shown in Plate 2.6. and was shown to be sensitive to RNase A and resistant to DNase (Plate 2.6., lanes 2-5). Some degradation of the RNA was seen, but high molecular weight RNA species were observed to be present by comparison with the DNA 1 kb ladder marker, ie. 8-10,000 base pairs.

2.3.4. COMPLEMENTARY DNA SYNTHESIS AND CLONING

The efficiencies and yields of the first and second strand cDNA reactions are calculated from the recovery of TCA precipitable counts after each reaction (Table 2.2.). The first strand reaction mixture contained 5 μg of heat denatured PVY^N RNA and [^{3}H]-dTTP at 1000 cpm/pmol. Total counts per minute (cpm) were calculated as 't_{60}'-'t_0' cpm.

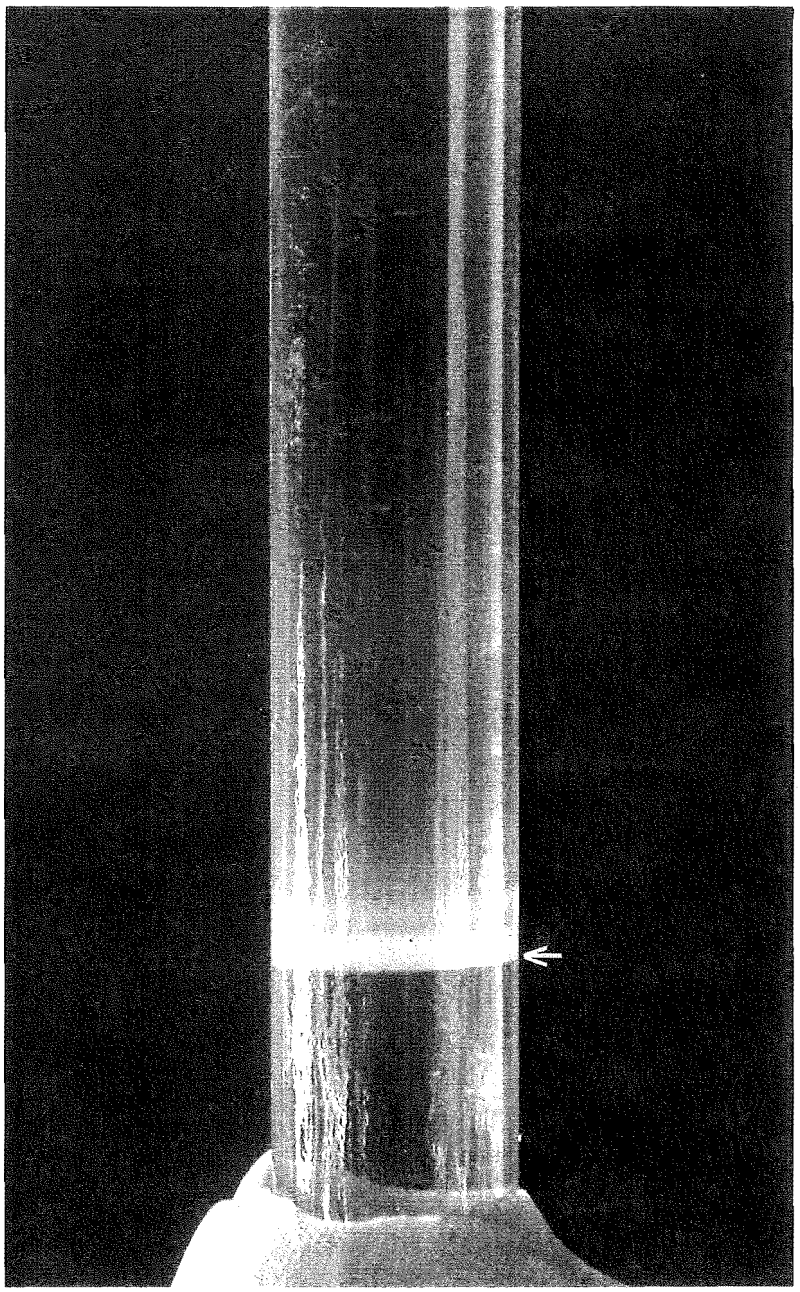


Plate 2.3. Caesium sulphate gradient after centrifugation at 30, 000 rpm for 16 hours in a Beckman SW55 rotor. The virus band (arrow) is separated from plant debris.

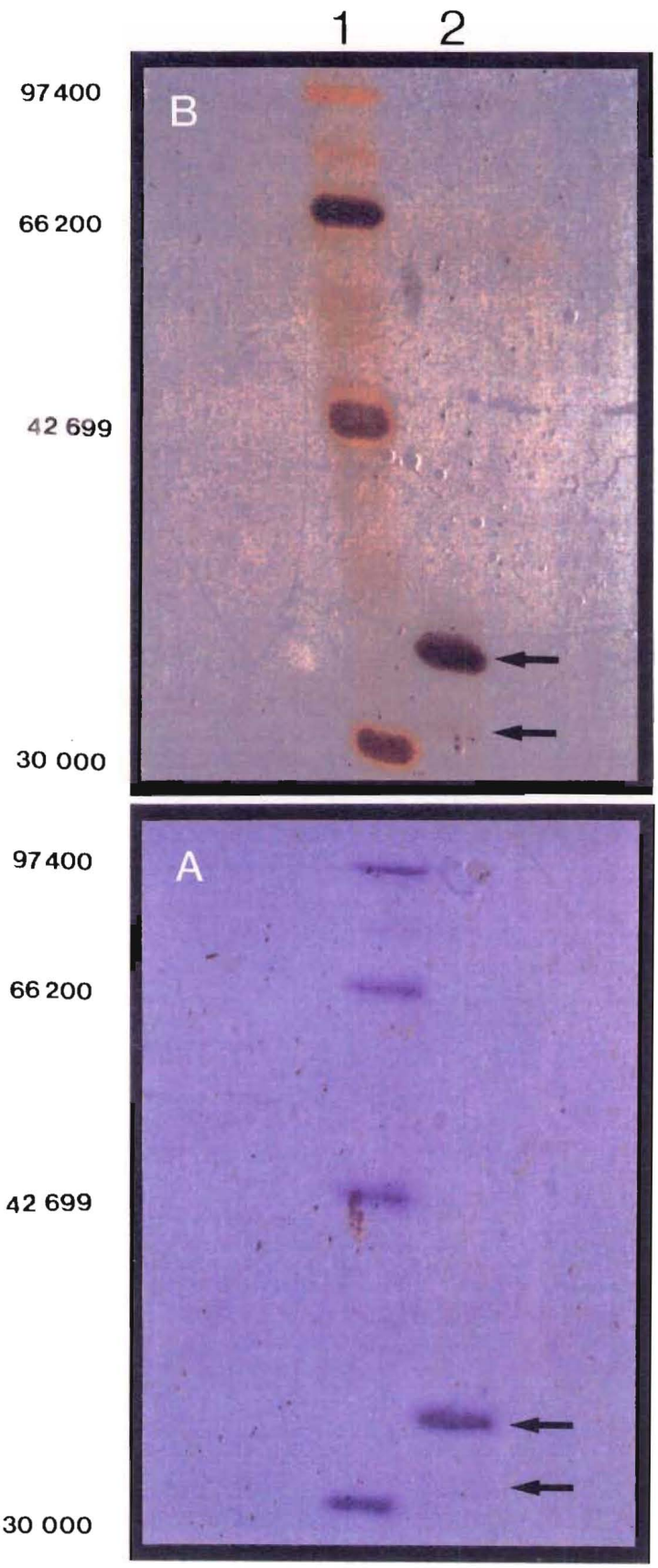


Plate 2.4. An SDS-PAGE on PVY^N. The gels are stained with **(A)** coomassie blue and **(B)** silver stain. In both gels **lane 1** contains protein markers (Biorad). **Lane 2** contains 1 μg of PVY^N whole virus. The two viral bands are indicated (34 kd and 31 kd).

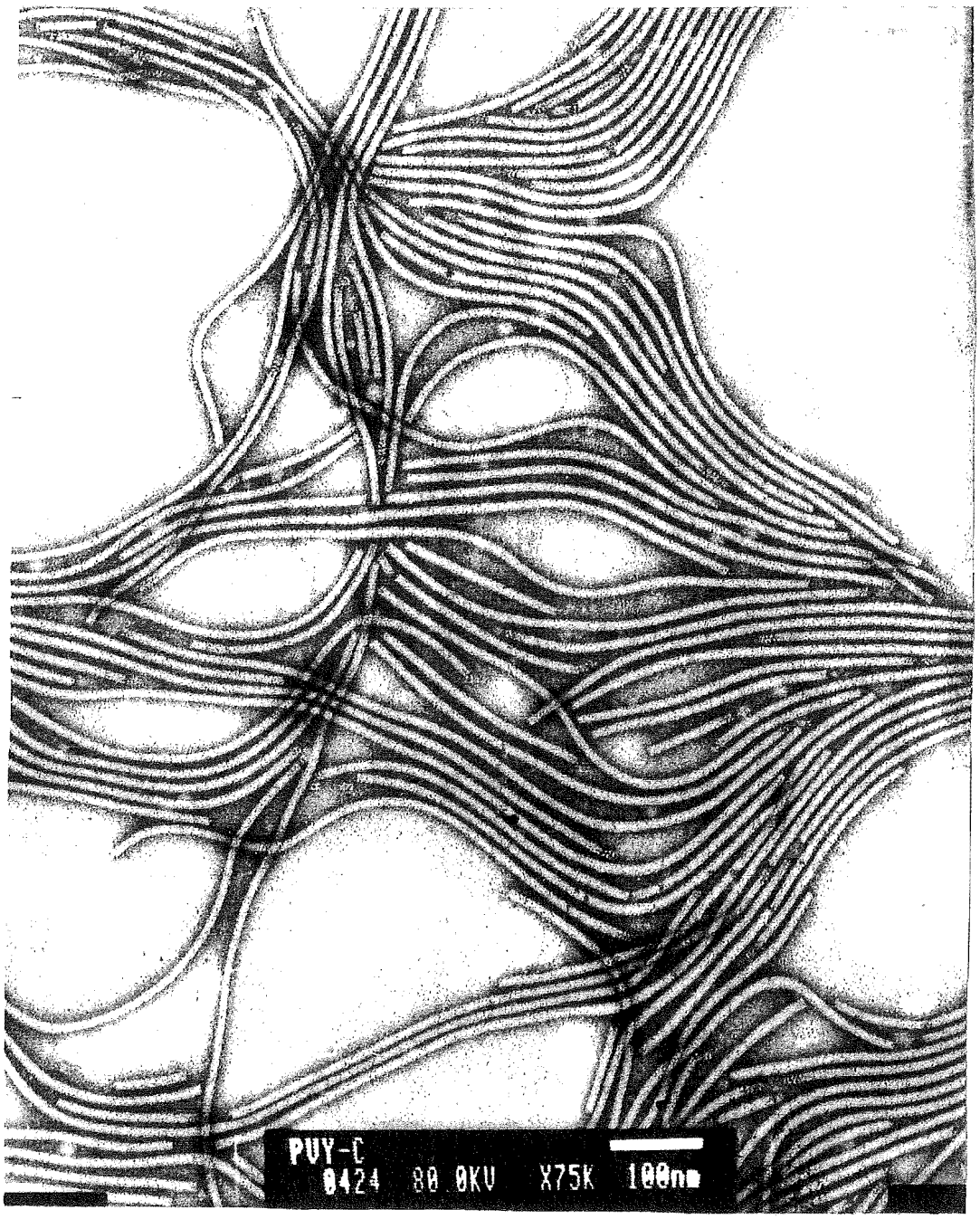


Plate 2.5. Electron micrograph of purified PVY^C virus particles negatively stained with phosphotungstic acid.

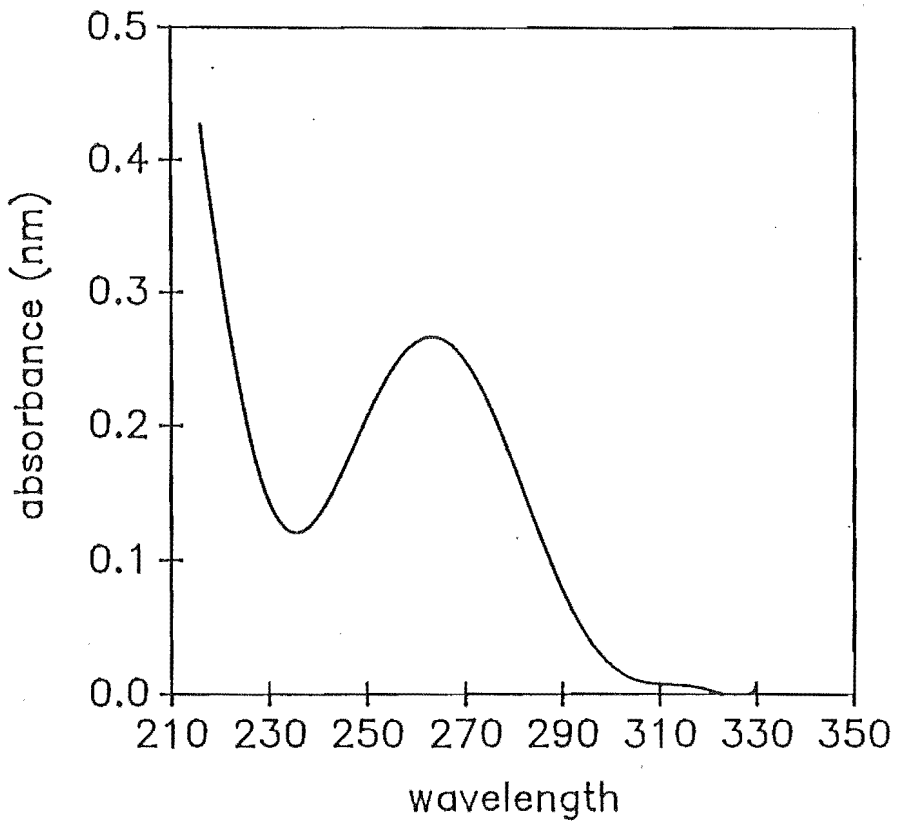


Figure 2.2. Ultraviolet absorbance spectrum of purified PVY^C whole virus.

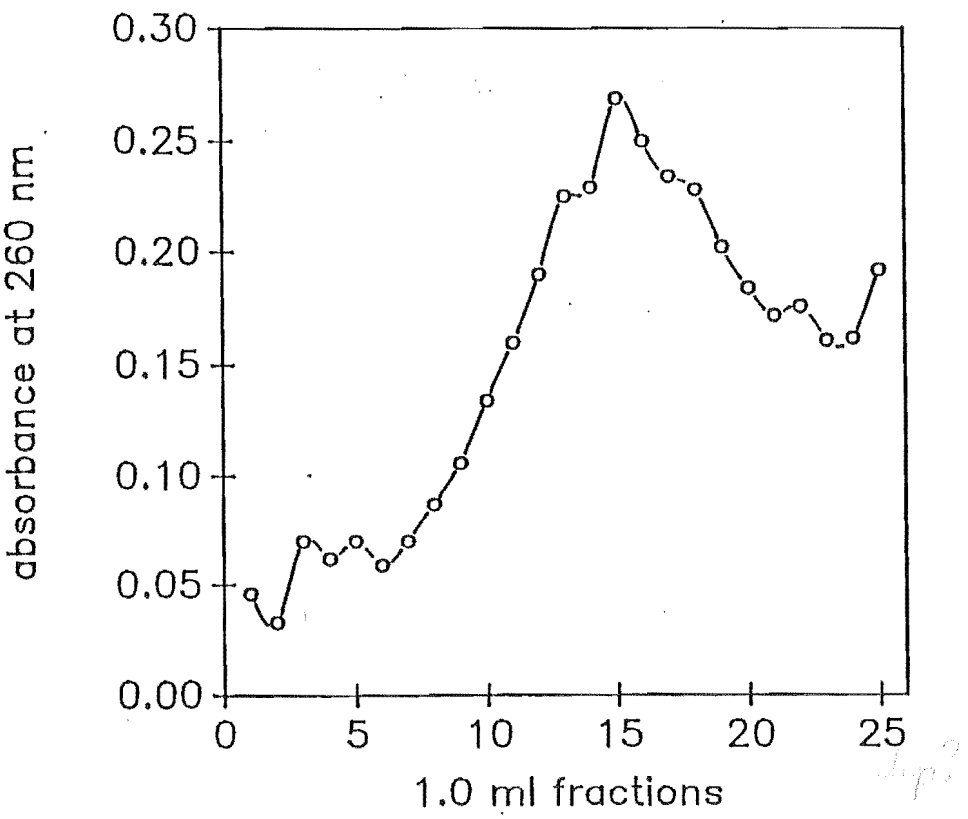


Figure 2.3. Absorbance at 260 nm of fractions collected from the sucrose gradient containing PVY^C RNA.

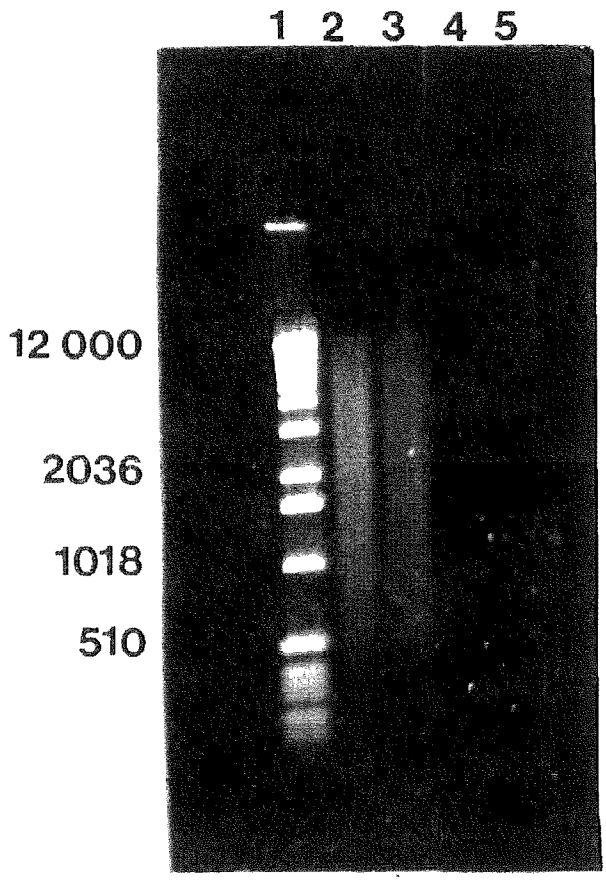


Plate 2.6. Electrophoresis of PVY^C RNA in a 1% agarose gel. **Lane 1.** BRL 1 kb DNA ladder. **2.** 1 μg PVY^C RNA treated with DNase. **3.** 0.5 μg PVY^C RNA treated with DNase. **4.** 1 μg PVY^C RNA treated with RNase. **5.** 0.5 μg PVY^C RNA treated with RNase.

Table 2.2. TCA precipitable cpms incorporated, cDNA yield and reaction efficiencies for first and second strand synthesis, and during cloning procedures.

Reaction	Total cpm	Yield	Efficiency
First strand	5000	260 ng	4%
Second strand	1000	253 ng	97%
Linker ligation:			
Spun G-50 column	2261	249 ng	98%
Ethanol precipitation	787.5	15.6 ng	6%

Double stranded DNA total cpm were calculated as 't_{120}'-'t_0' cpm, for [^{3}H]-dTTP at 250 cpm/pmol. Reactions containing either no M-MLV reverse transcriptase (in the first strand reaction) or *E. coli* DNA polymerase I (in the second strand reaction), and no oligo(dT) primer (in the first strand) were included as controls.

Table 2.3. Colonies produced on Amp$^+$, Xgal$^+$ plates after transformation of JM83.

Transforming DNA	Lac$^-$ (white)	Lac$^+$ (blue)
PVY^C cDNA/5'-OH pUC19[a]	14	30
Religated 5'-OH pUC19[b]	-	12
pUC19[c]	-	>2000
Untransformed JM83[d]	-	1

a. 100 ng 5'-dephosphorylated pUC19 ligated with 20 ng *Xba*I linkered PVY^C-cDNA
b. 100 ng religated 5'-dephosphorylated pUC19
c. 1 ng uncut pUC19
d. No DNA added

The efficiency of the first strand synthesis reaction was invariably low (approximately 5%) while the efficiency of the second strand reaction was usually between 90-100%. Losses of TCA precipitable counts during the cloning procedure are also summarized in Table 2.2.

The numbers of colonies resulting from transformation of PVY^C cDNA/pUC19 into *E. coli* strain JM83 are shown in Table 2.3. Fourteen recombinant plasmids were identified as being ampicillin resistant and unable to metabolise Xgal (lac$^-$). A good transformation efficiency of approximately 2000 colonies per ng of supercoiled plasmid was observed. An acceptably low background of blue colonies was also observed.

2.3.5. COLONY SCREENING

The fourteen lac- colonies were screened for the presence of inserts in the *Xba*I site of pUC19 (Plates 2.7. and 2.8.). Two colonies contained inserts of significant length: pVYC5 (2.4 kb) and pVYC11 (840 kb). The insert in pVYC5 was approximately the same length as pUC19 (2686 base pairs) and it was difficult to resolve the doublet on an agarose mini-gel (Plate 2.9., lane 6). To confirm the length of the insert a double digestion using endonucleases *Xba*I and *Sca*I was done. The *Xba*I cDNA insert was cleaved into two fragments of 1889 and 506 base pairs (bp) (Plate 2.9., lane 7). The pVYC11 insert appeared to be flanked by a single *Xba*I site (Plate 2.9., lane 8) and was resolved from the plasmid by a double digestion with *BamH*I and *Xba*I (Plate 2.9., lane 9). After re-streaking onto fresh ampicillin/X-gal LM plates, pVYC5 and pVYC11 were purified in large scale for further analysis.

2.3.6. DNA SLOT BLOT HYBRIDISATION

An autoradiograph showing hybridisation of a probe containing PVY^N coat protein sequences to selected plasmids is presented in Plate 2.10. The ^{32}P-labelled *Cla*I fragment of pVYN27 hybridised to pVYC11 and pVYN27, and to crude DNA prepared from tobacco plants infected with both PVY^C and PVY^N. No hybridisation was observed with pVYC5 or with controls pUC19, pBR322 and pGQ101. This indicates that the recombinant plasmid pVYC11 was of viral origin and contains PVY sequences derived from the coat protein gene, but that pVYC5 does not.

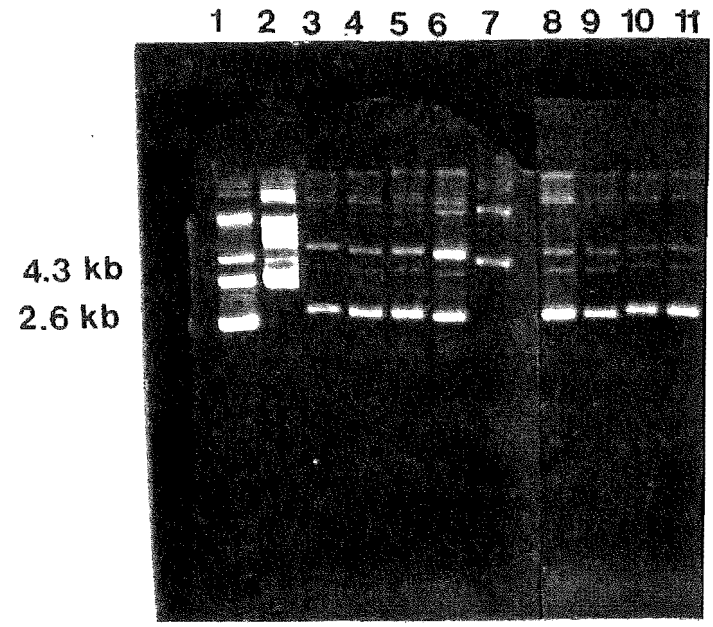


Plate 2.7. Electrophoresis of plasmids prepared from lac$^-$ colonies. **Lane 1.** 1 μg pUC19. **2.** 1 μg pBR322. **3-11.** pVYC1-pVYC9.

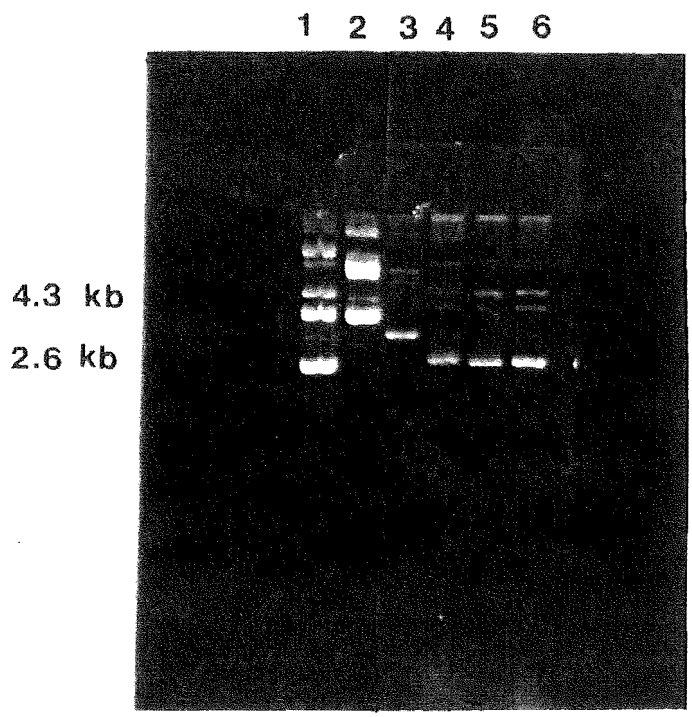


Plate 2.8. Electrophoresis of plasmids purified from lac$^-$ colonies. **Lane 1.** 0.7 μg pUC19. **2.** 0.7 μg pBR322. **3-6.** pVYC11-pVYC14.

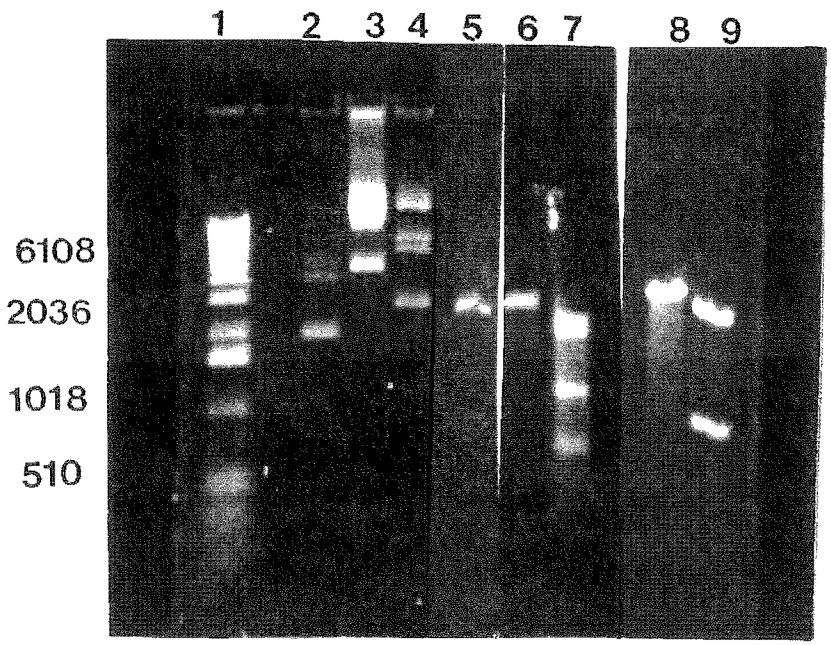
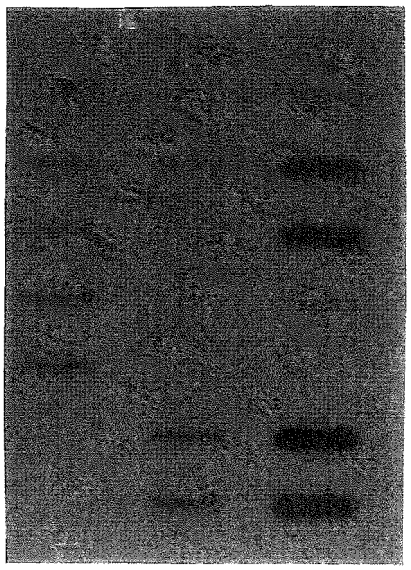


Plate 2.9. Electrophoresis of pUC19, pVYC5 and pVYC11 after digestion with selected restriction endonucleases. **Lane 1**. BRL 1 kb DNA ladder. **2**. Undigested pUC19. **3**. Undigested pVYC5. **4**. Undigested pVYC11. **5**. pUC19 digested with *Xba*I. **6**. pVYC5 digested with *Xba*I. **7**. pVYC5 digested with *Xba*I and *Sca*I. **8**. pVYC11 digested with *Xba*I. **9**. pVYC11 digested with *Xba*I and *BamH*I.



pVYC5	pUC19(2)	pGQ101
"	"	"
pVYC11	pUC19(3)	Infected PVY^N
"	"	"
pVYN27	pBR322	Uninfected
"	"	"
pUC19(1)	Infected PVY^C	*Cla*I fragment
"	"	"

Plate 2.10. **(A)** Analysis of selected plasmids probed with the PVYN27 cDNA insert. **(B)** Key to samples loaded on the membrane.

2.4. DISCUSSION

Strains PVY^N and PVY^C were propagated in the laboratory host *Nicotiana tabacum*. When inoculated with PVY^N, symptoms generally gave a good indication of potyvirus levels in the host plant. Only leaves with well developed veinal necrosis were included for virus purification. Leaves which were yellowing or without symptoms were omitted to reduce contamination by secondary metabolites, including phenolics, and to increase the yield of virus from the preparation. *N. tabacum* plants inoculated with PVY^C gave less easily defined symptoms. In comparison with PVY^N, the infection took longer to establish and yields were generally much lower, making this a more difficult strain with which to work. It was often necessary to test plants for PVY^C infection using the ELISA assay, to determine whether the virus concentration was suitable for purification.

Delgado-Sanchez and Grogan (1966) observed that directly inoculated leaves were a better source of virus than systemically infected tissues, and detected a decline in virus titre in the host plant sap 10 days after inoculation. However, ELISA assays performed on leaves which were either directly inoculated or systemically infected by PVY^N, revealed a similar concentration of virus after 8-10 days. Therefore, all leaves showing positive viral infection, either by ELISA or symptom development, were used for purification.

The major difficulties encountered in purifying large quantities of potyviruses can be attributed to their relatively low concentrations in infected sap (Edwardson, 1974), and to the tendency of virus particles to aggregate causing losses during low speed centrifugation steps (de Bokx and Huttinga, 1981). The highest yield of purified potyvirus reported was 40 mg kg^{-1} of leaf material ($M^cDonald$ *et al.*, 1976; Leiser and Richter, 1978 (in de Bokx and Huttinga, 1981)), but yields in the order of 2-10 mg kg^{-1} are more commonly observed (Stace-Smith and Tremaine, 1970).

Both aggregation and low virus concentrations were observed for all the purification schemes used in this study. Samples assayed by ELISA at consecutive steps during virus purification revealed significant amounts of virus remaining in the plant debris, particularly after the initial homogenisation and clarification steps. Stace-Smith and Tremaine (1970) commented that it was essential to suspend virus pellets in high molarity buffers for good recovery, for example, 0.5M potassium phosphate buffer pH 7.0, and observed that suspension in low molarity conditions (0.001M phosphate) caused large portions of virus to either aggregate or adhere to plant constituents. The success in purifying PVY using the methods of Reddick and Barnett (1983) and R. Forster (pers. comm.) confirmed that an initial extraction buffer with a high ionic strength resulted in much improved yields, compared with the procedures of Hamilton and Nichols (1978) or Dougherty and Hiebert (1980a). The most successful homogenisation buffer included the antioxidants ascorbic acid and SDDC, both of which Delgado-Sanchez and Grogan

(1966) suggested prevented discolouration of the homogenate and helped prevent aggregation by acting as chelating agents. Clarification of the supernatant with chloroform aided purification by partitioning the virus into the aqueous layer. Polyethylene glycol precipitations are a general method for concentrating biological macrostructures, and were useful in concentrating virus particles from aqueous phases. Standard density gradient procedures such as sucrose gradients, which differentiate particles on the basis of mass (and not density) proved inappropriate for concentrating PVY. Equilibrium centrifugation in Cs_2SO_4, however, successfully banded the virus, including aggregates as well as individual particles. Reddick and Barnett's procedure gave a clean virus preparation as indicated by inspection of the last high speed centrifugation pellet using TEM, and spectrophotometric data.

Two polypeptides were observed by SDS-PAGE. There have been a number of reports of heterogeneity of the coat protein by such analysis, including BYMV, PVY, TEV and pea mosaic virus. All are shown to migrate as two bands with the slower moving unit being 32-34 kDa and the faster being 26-28 kDa (Hiebert and M^{c}Donald, 1973; Hill *et al.*, 1973; Gough and Shukla, 1981). The probable reason for this is a partial degradation of the polypeptide by proteolytic enzymes. Recently, Shukla and Ward (1988a) suggested that degradation initially involved the removal of the amino- and carboxy- termini of the coat protein and that this may occur during storage at 4°C. Results for PVY^N and PVY^C support this observation.

RNA was isolated from whole virus by the method of Brakke and van Pelt (1970) using a buffer containing SDS which disrupted the virus, and proteinase K, which degraded proteins including ribonucleases. In experiments not presented, *in vitro* synthesised RNA labelled with [^{3}H]-UTP showed higher yields when purified from a sucrose gradient (31%) compared with phenol extraction and ethanol precipitation (12%). *In vitro* RNA purified by sucrose gradient centrifugation was also less degraded than that prepared by phenol/chloroform extraction (results not presented). Hence the final purification of PVY RNA was on a sucrose gradient. Hinostroza-Orihuela (1975) and Hari (1981), however, both reported infectivity of potyviral RNA after phenol/chloroform extraction. PVY^C RNA tended to sediment over several zones in the sucrose gradient (Fig. 2.3.) and smearing of the RNA on agarose gels indicated a number of degradation products. However high molecular weight RNA species (approximately 10,000 bp) were consistently observed. It was assumed that the quality of the virus particles prepared affected the quality of RNA recovered, and that broken or partially degraded virus particles contributed to the smaller RNA species seen on the gels. Also, the use of a non-denaturing gel system meant that secondary structure which alters the electrophoretic mobility of RNA species, was not abolished. A substantial degree of secondary structure has been indicated for other potyviral RNA, and Hill and Benner (1976) observed a decrease in electrophoretic mobility between 30-40% after formaldehyde treatment of a number of potyvirus RNAs.

Yields of PVY^N RNA were usually much higher than those of PVY^C. One possible reason for this is a 'critical mass' phenomenon. Large amounts of whole virus were required in order to

isolate any detectable RNA. Typically 5 mg or more of freshly purified virus were needed and this posed a problem when isolating PVY^C RNA. The disease produced by this strain never appeared to be as severe as that produced by PVY^N, and whole virus yields of PVY^C were consistently low.

Currently there is great interest in cloning cDNA copies of RNA virus genomes. Applications of recombinant technology such as 'genetically engineered cross-protection' have important implications in the applied biology of RNA viruses and their hosts. In this study purified RNA was used as a template for cDNA synthesis. The cDNA which was synthesised to PVY RNA, however, was not representative of the whole genome. It was assumed that oligo(dT) priming of first strand cDNA would increase the likelihood of cloning the 3'-terminal sequences of PVY RNA. As this study was interested in cloning the coat protein gene which has been shown to be encoded at the 3'-terminus in other potyviruses (Dougherty *et al.*, 1985; Allison *et al.*, 1985a), cloning of the whole PVY genome, although desirable, was not necessary. It was also assumed that oligo(dT) priming would reduce the likelihood of degraded or internal RNA species being used as templates for cDNA synthesis.

Fellowes (1988) observed that the labelling of reactions with radioactive nucleotides was only useful for estimating yields if the specific activity of the stock solution was not greatly diluted. 1000 cpm/pmol of [α-^{35}S]dCTP resulted in one labelled nucleotide being incorporated every 8000 nucleotides. This possibly contributed to the apparent low yield of first strand cDNA as determined by TCA precipitable counts, where a 10,000 nucleotide RNA template was used.

'Blunt ending' of PVY^C double stranded cDNA by T4 DNA polymerase I in preparation for the ligation of synthetic *Xba*I linkers was omitted on the assumption that *E. coli* DNA polymerase I used in the second strand synthesis, would produce a fairly large population of blunt ended molecules. A loss of 94% was observed between linker ligation and the final ethanol precipitation of the second strand cDNA. The final yield was 6% of the second strand initially synthesised, and the losses seemed primarily to be associated with the ethanol precipitation steps.

T. Turpen (in press) comments on a number of technical problems accentuated by cloning large RNAs. These include distinguishing cloning artifacts from authentic cDNA, for example, the ligation of two cDNA fragments into the vector, or failing to recover small internal restriction fragments which could complicate further analysis. Small changes due to sequence variation in the RNA population and error in first strand synthesis by the reverse transcriptase may also affect the biological activity of the product cDNA. Considerable difficulty is often encountered in obtaining terminal 5'-end sequence information. The production of many independent overlapping clones of each region, previously mapped by restriction enzymes, was one suggestion for overcoming the problem of sequence variation. Significantly shorter DNA inserts are often observed in recombinant colonies than the original size distribution of cDNA molecules synthesised. The ligation of cDNA into phage, rather than plasmid vectors, could possibly yield longer inserts.

Most of the recombinant plasmids screened for PVY^C cDNA inserts were less than 500 base pairs, and consequently were too short to encode the full coat protein gene sequence. The absence of longer cDNAs may have been due to the fragmentation of either the RNA template or the cDNA. Colonies containing the plasmids pVYC5 and pVYC11 were streaked out and rescreened for ampicillin resistance and β-galactosidase activity. Recombinants pVYC5 and pVYC11 contained inserts of approximately 2400 and 800 nucleotides, respectively.

One of the recombinant clones produced from PVY^N RNA by Fellowes (1988) was pVYN27. Hybridisation of a cDNA probe transcribed from PVY^N RNA to this recombinant plasmid and to PVY^N RNA, but not to non-recombinant plasmids, indicated that pVYN27 contained sequences from PVY^N. Subsequent sequence analysis of pVYN27 (Chapter 3.0) confirmed that the clone contained the 3'-sequence of the coat protein gene, and so this clone was used to probe recombinant plasmids pVYC5 and pVYC11. While pVYC11 gave a positive signal, pVYC5 did not. This suggested that pVYC5 did not contain PVY coat protein sequences but possibly sequences from elsewhere in the genome. This could not be discounted as internal *Xba*I sites were not protected by methylation prior to digestion with *Xba*I. Such sequences are rare in TEV and TVMV, and Turpen (in press) did not observe digestion of cDNA made to PVY RNA with *Xba*I. Also, as they occur approximately every 4^6 (4096) nucleotides, an *Xba*I site would be unlikely in this cDNA sequence. Further investigations into the origin of the pVYC5 inserted sequence were made by DNA sequencing (Chapter 3.0).

All further experiments reported in this study use the pVYN27 clone derived from PVY^N RNA. The clone from this strain was used because it encoded a longer cDNA sequence than pVYC11 and was more likely to encode the whole PVY coat protein gene; the PeMV coat protein gene is 1130 nucleotides long (Dougherty *et al.*, 1985). More significantly, PVY^N is the most recent PVY isolate found in New Zealand, it is a local isolate, and is highly aphid transmissable. PVY^N therefore poses a greater threat to current New Zealand potato and tobacco plantations, as well as pepper and tomato crops, and is consequently of greater agricultural economic value.

The sequences cloned to PVY^C could be used as primers for the production of cDNA to more 5' regions of the viral genome. The PVY^C clones would also be useful for a comparative study with PVY^N as the two strains are markedly different in their symptomology and host interactions.

CHAPTER THREE

DNA SEQUENCE ANALYSIS OF THE PVY^N COAT PROTEIN GENE

3.1. INTRODUCTION

3.1.1. GENERAL INTRODUCTION

The first part of this introduction discusses dideoxy DNA sequencing. The second section relates the application of DNA sequencing to potyviruses and outlines some of the results gained from this.

As early as 1970, knowledge of nucleic acid enzymology and chemistry was sufficiently advanced to anticipate the development of modern rapid methods of DNA sequencing (Hindley, 1983). The development of gel electrophoretic techniques, particularly the extraordinary resolving power of polyacrylamide gels run under denaturing conditions, and the application of recombinant DNA techniques and gene cloning to isolate, identify and amplify the DNA in question, have provided the means for DNA sequencing to become a reality (Hindley,1983). These methods also had implications for sequencing cDNAs produced by reverse transcription of RNA species and cloned into M13 (Hamlyn *et al.*, 1978; Houghton *et al*, 1980 (in Hindley, 1983)).

Currently there exist a plethora of sequencing methods, all of which have certain advantages. The majority can be divided into two broad groups for generating labelled oligonucleotides from which DNA sequences can be deduced: the primed synthesis approach and the chemical method. Both of these protocols depend on analytical polyacrylamide gel electrophoresis to resolve oligonucleotides which vary in length by a single nucleotide. The other prerequisite is the availability of microgram quantities of small defined fragments of DNA. Detection is via the incorporation of a radioactively or fluorescently labelled nucleotide residue into the DNA molecule which is to be sequenced. Two methods have established themselves as central to sequencing: the dideoxy chain terminator method of Sanger *et al.* (1977) and the chemical DNA sequencing method of Maxam and Gilbert (1977).

3.1.2. PRIMED SYNTHESIS DNA SEQUENCING

Sanger *et al.* (1973) described a primed synthesis method for sequencing two bacteriophage DNA sequences using DNA polymerase, primed with synthetic oligonucleotides. A DNA primer was hybridised to a single strand DNA template and the primer extended with the 'Klenow' fragment of DNA polymerase I (*E. coli*) in the presence of four deoxynucleotides (dNTPs), one of which was ^{32}P-labelled.

Sanger and Coulsen (1975) developed their procedure by fractionating the different sized cDNA fragments on high resolution denaturing polyacrylamide gels. This was the basis of the 'plus/minus' (+/-) method and the forerunner of all primer synthesis methods (Sanger and Coulsen, 1975). The approach reached a pinnacle with the determination of the entire 5375 nucleotide long DNA sequence of the phage ϕX174 (Sanger *et al.*, 1977). Additional back-up procedures were developed to confirm regions of deduced sequences. Depurination, a method devised by Burton and Petersen (1960), was a useful adjunct to the '+/-' method and confirmed the length and distribution of purine and pyrimidine tracts in fragments produced by primed synthesis. The single site ribosubstitution reaction (Brown, 1978 (in Hindley, 1983)) also added flexibility to the '+/-' system. A further mainstay of early DNA sequencing procedures was the 'wandering spot' method for analysing short DNA sequences (Sanger *et al.*, 1973).

Primed synthesis methods were revolutionised in 1977 by Sanger and his co-workers, with the introduction of chain terminating dideoxy nucleotides as specific inhibitors of DNA synthesis. Atkinson *et al.* (1969) showed that the inhibitory action of 2',3'-dideoxythymidine 5'- triphosphate (ddTTP) on DNA polymerase I depended on it replacing deoxyribothymidine (dT) in the growing oligonucleotide chain. DNA polymerase I requires a free 3'-hydroxyl group to initiate chain synthesis and facilitates the formation of a phosphodiester bond between the primer DNA 3'-hydroxyl and the 5'-phosphate of the substrate dNTP. As ddT has no 3'-hydroxyl group, the oligonucleotide chain being synthesised can not be extended beyond the point of ddT inclusion. Termination, therefore occurs specifically at this point. A primer/template hybrid incubated with DNA polymerase I in the presence of ddTTP, dTTP and the three other dNTPs produces a mixture of partially elongated fragments all possessing the same 5'-terminus and terminating in ddT at the 3'-end (Bankier *et al.*, 1987). Fractionation on a denaturing polyacrylamide gel generates bands corresponding to the series of fragments (Fig. 3.1.).

The incorporation of a dNTP or ddNTP into the extending DNA chain is a random event. By including analogues of dNTPs (ddNTPs) in the sequence reaction, a spectrum of primer extended fragments is generated. The relative frequency of chain termination events are varied by adjusting the relative concentrations of dNTPs and ddNTPs. Increasing the ratio of ddNTP to dNTP, increases the frequency of termination and produces shorter oligonucleotides. Typically a dNTP to ddNTP ratio of 1:100 gives a good distribution of bands (Hindley, 1983). The intensities of bands in different sequencing lanes can vary markedly as different ddNTPs do not compete for incorporation with the corresponding dNTP with equal efficiency (Hindley, 1983). For a given batch of dNTP and ddNTP the optimum ratios for giving bands of good intensity on a gel must be determined empirically. An alternative to ddNTPs as analogues are the arabinoside triphosphates (Hindley, 1983). Arabinose is a stereoisomer of ribose and a chain synthesis inhibitor comparable to the ddNTPs.

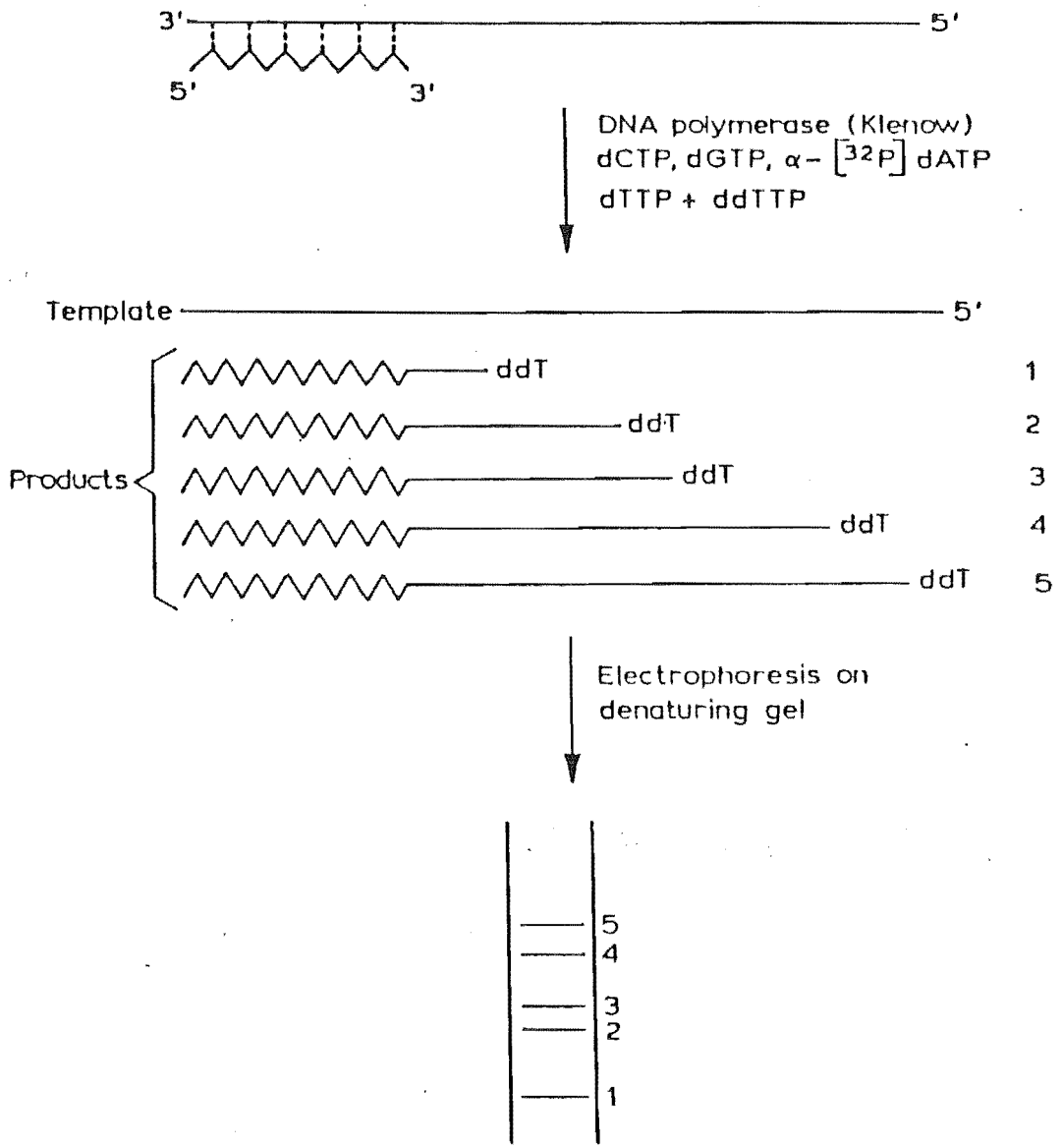


Figure 3.1. The principle of the dideoxynucleotide chain-termination method for DNA sequencing (after J. Hindley, 1983).

A common feature of primer synthesis methods is the need to reduce the large molecular weight DNA to be sequenced into sets of fragments. These can then be separated, purified and cloned into vectors for selection and amplification of the recombinant sequences. This provides the raw material for sequencing and hence recombinant DNA techniques have become inextricable from sequencing procedures (Bankier *et al.*, 1987). Plasmids and phages replicate their DNA independently of their bacterial host chromosome. Within the DNA of these vectors are sequences, either between genes or within non-essential genes, into which foreign DNA sequences can be introduced. DNA which has been fragmented by restriction endonucleases can be cloned into the replicative form (RF) of a natural ssDNA phage, for example M13 or fd phage, to produce recombinants which also contain the cloned sequence in single stranded form (Hindley, 1983). Amplification in M13 and its derivatives yields single stranded recombinant DNA which can be sequenced directly without further excision and fragment purification. These serve as the templates required for the primed synthesis reaction.

Sanger *et al.* (1980) reported the random, or 'shotgun', cloning of DNA fragments into the RF of the vector M13mp2 (Gronenbom and Messing, 1978) and used this approach in the analysis of the human mitochondrial DNA sequence. M13mp2 is a vector designed by Messing *et al.* (1977) and was constructed by the insertion of a restriction fragment of the *E. coli lac* regulatory region into wild type M13. The inserted fragment contained the *lac*-I gene, the *lac* promotor, operator and the proximal region of the *lac*-Z gene which encodes the first 145 amino acids of the α-peptide of the β-galactosidase gene. There exist a number of derivatives of this vector, for example M13mp7 (Messing *et al.*, 1981). The flexibility of the system was further improved by the introduction of M13mp8 and mp9 (Messing and Viera, 1982). These vectors contain a multi-purpose cloning site coding for eight discrete restriction endonuclease recognition sites, each in an alternate orientation. They therefore facilitate asymmetric cloning; inserted fragments in M13mp8 were automatically 'turned around' when recloned into M13mp9. When these phages infect a derivative strain of *E. coli* lacking the *lac pro* region, complementation occurs and a functional β-galactosidase is produced. The detection of plaques with DNA inserted in the cloning site, relies on the inactivation of β-galactosidase activity in the presence of Xgal (section 2.1.2.). Other features make M13 phages applicable for sequencing procedures. These include their small size (7200 base pairs), high copy number (200-300 molecules of supercoiled DNA (RF) per cell), and the ease of isolating the RF from infected bacterial cells. An alternative to M13 as a cloning vector is the phage fd tailored by Herrmann *et al.* (1980). Sites for insertion are usually located within genes coding for antibiotic resistance and this provides a marker for the selection of recombinants.

Fragments generated by restriction enzymes tend not to give a truly random array of DNA sequences (Hindley, 1983). There has been a marked drift toward the use of less directed degradation methods to obtain a random pool of DNA fragments for cloning. These alternatives

include DNA endogenases (DNase I and II), shearing by sonication and digestion with exonuclease III (Bankier *et al.*, 1987).

The initial requirement for primed synthesis sequencing is the formation of a primer/template hybrid. Previously, primers were restriction fragments isolated from template DNA. Alternatively, a variety of recombinant templates can be sequenced using a single 'universal' primer; fragments isolated from RF M13 (Heidecker *et al.*, 1980) or products of exonuclease III activity (Smith, 1979). These hybridise specifically to a region of ssDNA within the vector which is adjacent to the cloned insert. Various abbreviated primers have been developed and synthetic primers are now more commonly used, for example Messing *et al.* (1981). Synthetic universal M13 primers, 15 and 17 nucleotides long, are commercially available. Other examples include a 26 base pair primer isolated from phage pSP14 (Anderson *et al.*, 1980 (in Hindley, 1983)). Priming from sets of internal primers derived from the original duplex DNA (Hindley and Phear, 1981) and from DNA produced by random shearing are also described.

In general, Sanger's chain termination sequencing method (Sanger *et al.*, 1977, 1980) allows up to 200 nucleotides from the priming site to be determined with reasonable accuracy. It is generally applicable to any DNA that can be obtained in single stranded form, and has probably been the method of choice for DNA sequencing among primed synthesis techniques (Hindley, 1983). The major limitation of the method is the resolving power of the gel system and the restriction on lengths of DNA, approximately 2 kb, which can be cloned into M13. Artifacts are occasionally observed and problems due to the 'pile up' of bands occur. Primed synthesis methods generally do not reveal the presence of methylated or modified bases, as they analyse DNA via its complementarity and can only give sequence information in terms of the four canonical bases (Hindley, 1983).

An important feature of Sanger's method is that it can easily be modified to gather large amounts of data for large sequencing projects. These are reviewed by Barnes (1987). Aside from the shot-gun sub-cloning of DNA in M13 vectors which has greatly accelerated the sequencing of long DNA fragments, many modifications have been made. Some examples include the development of the new series of 'pEMBL' plasmids. These are useful for preparing large quantities of plasmid and for the stability of long cDNA inserts (Dente and Cortese, 1987). The use of transposon promoted deletions of DNA allows a directed approach for generating overlapping DNA sub-fragments (Ahmed, 1987). Considerable efforts have been directed toward improving electrophoretic methods with field strength gradients, wedge shaped gels and direct transfer electrophoresis (Pohl and Beck, 1987). Primers labelled at the 5'-end by one or four different fluorescent dyes have successfully been used for direct laser-supported determination of sequence during gel electrophoresis (Smith *et al.*, 1987).

Computer programmes have also developed which can collate and evaluate the huge data sets generated by DNA sequencing (Staden, 1986).

3.1.3. SEQUENCING OF POTYVIRUSES

To date the potyvirus group numbers 152 definitive and possible members (Francki *et al.*, 1985). It contains an unusually large number of strains which differ mainly in their biological properties, for example host range. Viruses have been assigned to the potyvirus group on the basis of particle length, physical properties, transmissibility, host range and serology (Harrison *et al.*, 1971).

It has consistently been pointed out that the taxonomy of the potyvirus group is unsatisfactory and that the resolution of potyvirus taxonomy presents a major challenge to plant virologists (Francki, 1983; Francki *et al.*, 1985; Harrison, 1985; Shukla and Ward, 1988b). This is due to the large size of the group, the vast variation among members and the lack of taxonomic parameters by which to distinguish its distinct members (Francki *et al.*, 1985). Reclassification can be achieved only from a thorough understanding of the genomic variation, sequences and structure of potyviruses.

Nucleic acid hybridisation is potentially useful for establishing the identity of potyviral isolates. However, the nucleotide sequences of the genomes of distinct potyviruses sequenced to date show little extensive homology. The 3'-untranslated region of potyviral RNA nucleotide sequences have been observed to differ considerably in length and to display no significant sequence similarity between distinct potyviruses (Frenkel *et al.*, 1989). By contrast, the 3'-untranslated sequence in related strains is similar in length and sequence, and cDNA probes to this region could be useful for detecting strains of a potyvirus. Antibodies to selected regions of proteins may be a more realistic method for identifying potyviral isolates. A computer search for identical peptide regions, seven or more amino acids long, revealed 41 matching regions in the polyproteins encoded by TEV (Domier *et al.*, 1986) and TVMV (Allison *et al.*, 1986). Shukla suggested that coat protein sequence data could be used as a solid basis for discriminating independent members of the group from related strains (Shukla and Ward, 1988a).

Primary, Secondary and Tertiary Structure of Coat Protein Sequences

The complete genomic sequence is known for two potyviruses, tobacco vein mottling virus (TVMV) (Allison *et al.*, 1986) and tobacco etch vius (TEV) (Domier *et al.*, 1986). As for the picorna- and comoviruses, potyvirus genomes appear to be translated as one large polyprotein precursor. In addition to TEV and TVMV, the primary structure of the coat proteins, determined either by gene or protein sequence, of nine other distinct potyviruses (making a total of 25 strains) are known (Shukla and Ward, 1988a).

Complete amino acid sequences have been obtained for the coat proteins of PVY^{D} (Shukla *et al.*, 1986), Johnson grass mosaic virus (JGMV) (Shukla *et al.*, 1987) and passionfruit woodiness virus (PWV) (Shukla *et al.*, 1988b). Partial peptide sequence data are available for sugarcane mosaic virus (SCMV) (Shukla *et al.*, 1987). Gene sequencing has predicted coat proteins sequences for plum pox virus (PPV) (Ravelonandro *et al.*, 1988), pepper mosaic virus

(PeMV) (Dougherty *et al.,* 1985), soybean mosaic virus, N-strain (SMV-N) (Eggenberger *et al*, 1988) and V-strain (SMV-V) (Gunyuzlu *et al.*, 1987), PVY^{N*} (Dutch isolate) (van der Vlugt *et al.*, 1989), TVMV and TEV.

Attempts to predict potyviral coat protein tertiary structure by various algorithms have generally been unsuccessful (Shukla and Ward, 1988a). However these methods have some value in indicating local regions of secondary structure and for detecting similarity in secondary structure between related proteins. Structural predictions for PVY coat protein using the method of Chou and Fasman (1974) indicated ten regions of α-helix and four sections of β-sheet (Shukla and Ward, 1988a). They observed a number of parallel features occurring in other rod-shaped viruses, for example, tobacco mosaic virus (TMV) (Bloomer *et al.*, 1978) and potato virus X (PVX) (Sawyer *et al.*, 1987), including four extensive α-helical sections. A number of conserved arginine residues are present in potyvirus coat proteins and may have an interactive role with viral nucleic acids as suggested for tobamoviruses and potexviruses. The primary structure data available for potyviruses show only one cysteine residue which is consistently conserved (Shukla and Ward, 1988a). They also observed that distinct potyviruses differed in their number of cysteine residues but no investigations have been made to determine the role of these in interchain disulfide bonding or interchain disufide dimer formation.

Apart from limited studies on the assembly of the PVY coat protein into particles (M^{C}Donald *et al.*, 1976; M^{C}Donald and Bancroft, 1977), little is known about the three dimensional structure of the coat protein monomer and its packing arrangements in the virus particle. It has recently been shown (Allison *et al.,* 1985a; Shukla *et al.*, 1988c; Dougherty *et al.*,1985) that the amino and carboxy (N- and C-) terminal regions of potyvirus coat proteins are exposed on the virus particle surface. The C-termini are highly homologous (approximately 65% identity). The N-termini from distinct potyviruses, however, differ markedly in length and sequence while strains of one virus are highly homologous in this region.

Mild proteolysis of the coat protein with trypsin removed 30 or more N-terminal residues and 18-20 C-terminal amino acids, leaving a 'core' protein of 215-218 amino acid residues (Shukla *et al.*, 1988c). These core particles were indistinguishable from untreated native particles when examined under the electron microscope and were still infectious. The role of the surface exposed N-and C-termini of the coat protein in structural or biological functions of potyviruses is not known, and their removal does not prevent the remaining core coat protein from forming rods without RNA. Particles of three other rod-shaped plant groups, potex-, tobamo- and tobraviruses, also have coat protein N- and C- termini exposed on external surfaces and this suggests the possibility of common features for polypeptide folding and subunit packing (Shukla *et al.,* 1988c).

Serology

Serology has been a criterion for classifying members of other plant virus groups, for example the tobamoviruses (Gibbs, 1977) and the tymoviruses (Koenig and Leseman, 1979). To date it has been unsatisfactory when applied to the large potyvirus group. Serological relationships between related strains and distinct members have been found to be complex and inconsistent (Hollings and Brunt,1981; Francki, 1983; Moghul and Francki, 1976).

Different methods for preparing antisera to potyviruses have been shown to produce probes which differ in specificity and quality. There are a number of variables which contribute to this variation, such as the method of antigen purification, the time of bleeding to obtain the antibodies and the number of bleeds. Due to these factors, different antisera are not comparable. Also, the use of broad spectrum polyclonal probes has been compromised by the observations of apparently inconsistent serological relationships between pairs of distinct potyviruses, for example JGMV (*Gramminacae* hosts) and WMMV-II (*Curcurbitacae* hosts) (Shukla *et al.*, 1988c). The significance of these pairings between biologically unrelated potyviruses can not be assessed until the epitopes responsible have been mapped, and it is not clear whether these paired relationships have any value in potyvirus taxonomy (Shukla and Ward, 1988a). A final problem in potyviral serology is posed by viruses like PeMV and PVY. These have been reassigned as strains of the same virus on the basis of very similar coat protein sequence, but have only a distant serological relationship. Hence the difficulty with the previous classification of potyviruses. Novel approaches to the serology of potyviruses have been advanced by Shukla based on the concept of virus specific epitopes being located in the N-terminus region of the potyvirus coat protein (Shukla *et al.*, 1988c). The N-terminus constitutes the most immunodominant region in the potyvirus. It is the only large region in the entire coat protein that is highly variable and so virus specific epitopes within this region generate virus specific antibodies. Using these as diagnostic probes, strains of a distinct virus can be distinguished. However such probes are restricted to viruses whose coat protein N-terminal sequences are known. Antibodies raised against the trypsin dissociated 'core' protein have recognised all 15 definitive potyviruses tested so far and therefore represent broad spectrum probes for detecting potyviruses (Shukla and Ward, 1988a).

Taxonomy

Identification and subgroupings of the potyviruses based on host range, cross-protection and morphology of cytoplasmic inclusions, show anomalies and inconsistencies, and therefore have limitations in classifying the potyviral members (Hollings and Brunt, 1981; Francki *et al.*, 1985). Symptomology and host range have played a significant role in the delineation of potyviruses and their strains in the past. However reliance on these criteria has created confusion as different potyviruses have been shown to produce different symptoms in hosts depending on environmental conditions and cultivars of plant species (Shukla and Ward, 1989). The

characteristic 'pinwheel' type inclusions found in host cells infected with potyviruses have previously been used as diagnostic markers for establishing hierarchical classification of potyviruses (Edwardson *et al.*, 1974). Given the need to revise the virus/strain status of the group, the value of cytoplasmic inclusions as a taxonomic tool is questionable (Shukla and Ward, 1989). The generally accepted idea that related strains of the same virus are capable of cross-protection, whereas distinct viruses are not, may prove useful as a taxonomic criterion once the placement of viruses and strains has been reassigned (Shukla and Ward, 1989). Examples of unexpected cross-protection between viruses hitherto considered distinct indicate a need for caution when assigning classification on the basis of cross-protection experiments.

Shukla and Ward, (1988b) advocate the structural properties of coat proteins as useful for the identification and classification of group members for the following reasons:

1. the coat protein is a unique gene product whose amino acid composition is characteristic of the group. In contrast to other gene products, it shows no significant homology with other groups of plant viruses (Domier *et al.*, 1987).
2. the coat protein accounts for 95% of the potyvirus virion mass, although it is encoded on only 12% of the genome.
3. the serology reflects the protein structure, and serological techniques are the most preferred to date for identifying and classifying plant viruses.

Shukla's group have developed a method for peptide profiling potyvirus coat proteins using high performance liquid chromatography (HPLC) (Shukla *et al.*, 1988a). Tryptic digests of the coat proteins of four strains of PVY, two strains of PWV, three strains of SCMV and one strain of bean yellow mosaic virus (BYMV), JGMV and watermelon mosaic virus (WMMV-II), showed that peptide patterns of strains of the same virus were very similar, but those from distinct potyviruses were quite different. The peptide profiles were highly reproducible and suggested a useful method for differentiating potyviruses and strains. Profiling of N-terminal peptides after trypsic proteolysis by HPLC revealed far simpler diagnostic patterns than the complete peptide digests and should be sufficient to diagnose distinct viruses and strains.

Currently, coat protein structural information provides a sound basis for identification and classification of potyviruses. Computer analysis of the nucleotide sequences of 20 strains of 9 viruses (Shukla and Ward, 1988b), showed that distinct members exhibited sequence homologies of 38-71% (average 54%) and strains of individual viruses have sequence similarities of 90-99% (average 95%). These results clearly reveal a bimodal distribution and a clear demarcation of sequence similarity between distinct viruses and strains. This casts doubt on the 'continuum' hypothesis previously proposed to explain the unsatisfactory taxonomy of the group. The hypothesis states that potyviral strains form a continuum array so that the boundaries which separate strains of distinct viruses cannot be sharply defined (Bos, 1970; Harrison, 1985). Some potyviruses have been redefined and classified as a result of this bimodal distribution. In one

example, JGMV was found to be an independent member of the group (Shukla *et al.*, 1987) and not a strain of SCMV as was previously believed.

In conclusion, it appears that classical methods used in past identification and classification do not provide suitable criteria for distinguishing distinct potyviruses from strains, or for developing a taxonomy. Immunochemical analysis of native particles, core protein and trypsin treated protein have provided an insight into the molecular basis for potyviral serology and help explain some problems associated with conventional serology. Sequence data from whole genomes should provide the ultimate criteria for classification. Coat protein sequence data appear to be a good index for genetic relatedness, and are convenient for developing a taxonomy of the potyviruses (Shukla and Ward, 1989).

3.2. MATERIALS AND METHODS

3.2.1. FRAGMENT ISOLATION

Table 3.1. Restriction endonucleases and incubation conditions

R.E.[a]	Buffer (10 x)[b]	Temperature
*Xba*I	50mM Tris-HCl pH 8.0, 10mM $MgCl_2$, 50mM NaCl	37^{o}
*Hae*III	50mM Tris-HCl pH 8.0, 10mM $MgCl_2$, 50mM NaCl	37^{o}
*Taq*I	50mM Tris-HCl pH 8.0, 10mM $MgCl_2$, 50mM NaCl	65^{o}
*Rsa*I	50mM Tris-HCl pH 8.0, 10mM $MgCl_2$	37^{o}
*Acc*I	50mM Tris-HCL pH 8.0, 10mM $MgCl_2$	37^{o}
*Alu*I	50mM Tris-HCl pH 8.0, 10mM $MgCl_2$	37^{o}
*Hpa*II	10mM KCl, 10mM Tris-HCl pH 7.4, 10mM $MgCl_2$, 1mM DTE, 100 μg/ml acetylated BSA	37^{o}
*Sma*I	20mM Tris-HCl pH 7.4, 5mM $MgCl_2$, 50mM KCl	37^{o}

a. Restriction endonucleases (R.E.) (except for *Hpa*II) supplied by BRL. *Hpa*II supplied by NEB.

b. 10 x buffer supplied by BRL.

Twenty μg of plasmid pVYN27 were cleaved with 20 units each of the following restriction endonucleases in 150 μl of the appropriate buffer: *Xba*I, *Hae*III, *Taq*I, and *Rsa*I (Table 3.1.). After one to two hours digestion, 45 μl of 5x sample buffer (50% glycerol, 0.25% bromophenol blue in 5x TBE) were added to the reaction mix and 20 μl (approximately 2 μg of DNA) were loaded onto nine lanes of a 6% polyacrylamide gel (6% acrylamide, 1x TBE, 0.1% ammonium persulphate, 15 μl TEMED). One μg of pUC19 was digested with the appropriate restriction endonuclease in a 20 μl volume and was loaded onto a tenth lane to indicate which fragments were plasmid bands. The gel was electrophoresed in 1x TBE at 90 V for three hours or until the blue dye front had migrated 16 cm, then lightly stained with a 0.5 mg ml^{-1} solution of ethidium bromide, and viewed on a short wavelength ultraviolet transilluminator to visualise the bands. Each insert fragment band of pVYN27 was excised from the gel and pooled in an eppendorf tube. The DNA was eluted from the crushed gel slices in 0.5-1.0 ml of 0.5M NH_4 acetate, 1mM EDTA, overnight at $37^{o}C$.

The mixture was spun in a microfuge at 13,000 rpm for 10 minutes and the supernatant drawn off and collected into a clean eppendorf tube. A further aliquot of extraction buffer was

added to the crushed gel slices, vortexed, spun and the supernatants collected. Two volumes of 100% ethanol were added to the pooled supernatants. The fragment DNA was precipitated at -80^{o}C for 10 minutes and pelleted at 13,000 rpm for 30 minutes at 4^{o}C. The visible pellets were dried and resuspended in 5-10 μl TE pH 8.0.

One μl aliquots of the isolated fragments were electrophoresed on a 1% agarose mini gel in 1x TBE to determine the purity of each fragment. The approximate concentration of the DNA was calculated by comparing the intensity of fluorescence with the bands generated from 1 μg of molecular weight DNA marker (BRL).

Alternately, sub-fragments of insert were generated which could be cloned straight into the vector without polyacrylamide gel purification. Five μg of pVYN27 insert were digested in 20 μl of appropriate buffer with the following endonucleases: *Taq*I, *Hpa*II and *Alu*I (Table 3.1.). After one to two hours digestion, the mix was phenol extracted and the DNA precipitated in ethanol. The dried DNA pellet was resuspended in 10 μl TE pH 8.0, in preparation for ligation.

3.2.2. SUB-CLONING INSERT FRAGMENTS INTO A SEQUENCING VECTOR (BANKIER *et al.*, 1987)

Preparation of Vector for Ligation

The replicative forms M13mp18 and M13mp19 (donated by D. Hill, Department of Biochemistry, University of Otago, Dunedin, New Zealand) were made linear by digestion with either *Sma*I (for ligation of blunt ended fragments made by *Rsa*I, *Alu*I and *Hae*III digestion), *Xba*I (for ligation with *Xba*I generated inserts) or *Acc*I (for ligation with *Taq*I and *Hpa*II generated fragments).

Five μg of M13mp vector were digested with the appropriate endonuclease (Table 3.1.) in a 50 μl volume. Complete digestion was confirmed by electrophoresing 5 μl aliquots on a 1% agarose mini gel in 1x TBE. The reaction mixture was then phenol/chloroform extracted and the DNA was ethanol precipitated. To prevent recircularisation of the vector and thereby reduce the background of non-recombinant M13 plaques, the linear vector was treated to remove 5'-phosphate groups. The linear vector DNA was resuspended in 5 μl 10mM TE pH 8.0 and added to 45 μl of CIP buffer and 0.1 μl of CIP enzyme (section 2.2.5.). The vector DNA pellet was dried, resuspended in 10mM TE pH 7.5 to 20 ng μl^{-1} and stored in aliquots at -20^{o}C.

Ligation of Vector and Fragment DNA

Vector M13mp18 DNA and fragmented insert DNA were ligated at 16^{o}C overnight in a 10 μl ligation reaction (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 10mM DTE, 1mM spermidine, 1mM ATP, 100 μg ml^{-1} BSA) with 1 μl T4 DNA ligase. 40 ng of M13 vector and 10 to 100 ng of isolated fragment were used in each ligation. The following two ligation controls were included: first, self-ligated vector containing no insert fragment reflected the efficiency of the dephosphorylation reaction, and second, ligation of vector with 10 ng of *Alu*I cut pBR322 was a positive control for the ligation reaction.

Transformation

E. coli strain TG1 cells were made competent and transformed with the M13/fragment ligation reactions, using a modification of the $CaCl_2$ method of Mandel and Higa (1970). Ten ml of 2x YT media (1% bactotryptone, 1% yeast extract, 0.5% NaCl) were inoculated with a single colony of TG1 cells and agitated overnight at 37°C. One ml of the overnight cell suspension was used to inoculate 100 ml of 2x YT media and shaken at 37°C until the cell suspension had an A_{550} of 0.45-0.55. The cells were collected by centrifugation at 2500 rpm for five minutes in a chilled centrifuge tube. The pelleted cells were gently resuspended in cold 50mM $CaCl_2$ to half their original volume and incubated on ice for thirty minutes. The cells were pelleted as before, resuspended in cold 50mM $CaCl_2$ to 1/12.5 of their original volume and stored in capped tubes at 4°C.

Two hundred microlitres of competent TG1 cells and 10 μl of ligation mixture were added to a sterile glass tube and incubated on ice for 45 minutes. The mixture was then heat shocked at 42°C for two minutes and added to 3.0 ml of top agar (1% bactotryptone, 0.8% NaCl, 0.8% bactoagar) containing 30 μl of 2% Xgal in dimethylformamide, 20 μl of 2.5% IPTG and 200 μl of an overnight culture of TG1 cells, held at 45°C. This agar mixture was immediately poured onto TYE plates (2x YT, 1.5% agar), spread to distribute the cells evenly, and allowed to set at room temperature for 15 minutes. The plates were inverted and incubated at 37°C overnight to establish plaques.

3.2.3. TEMPLATE PREPARATION (BANKIER *et al.*, 1987)

Each colourless plaque arose as a result of transformation by a recombinant M13 RF. The phage isolated from a single plaque were used to prepare single stranded recombinant phage DNA.

Ten ml of 2x YT media were inoculated with a single colony of TG1 cells and grown with shaking overnight at 37°C. The overnight cell culture was diluted 100x with 2x YT, and 1.5 ml of diluted culture were dispensed into sterile culture tubes. A separate single plaque was toothpicked into each tube and grown with vigorous shaking at 37°C for 4.5-5 hours. The cell suspensions were then transferred to eppendorf tubes and microfuged at 13,000 rpm for five minutes to pellet the cells. The supernatants were transferred to clean eppendorf tubes and 200 μl of PEG solution (20% PEG, 14.6% NaCl) were added to each tube. The phage pellet was collected by microfuging at 13,000 rpm for five minutes and the supernatant discarded. The pellet was re-spun to remove any residual PEG. The pellet was then resuspended in 100 μl TE pH 8.0 and extracted with an equal volume of buffer-saturated phenol. The aqueous phases were separated by microfugation and DNA precipitated from the aqueous phases in 70% ethanol and 0.3M Na acetate at -80°C for 10 minutes. The DNA was pelleted, washed with 1.0 ml of 100% ethanol, dried and resuspended in 30 μl TE pH 8.0.

3.2.4. SEQUENCING POLYACRYLAMIDE GELS

For sequencing uncharacterised clones, two loadings on 6% polyacrylamide, 8M urea gels (Sanger and Coulson, 1978) were made. The gels were electrophoresed for 1.5 and 4 hours respectively.

Typically a 50 cm gel and 0.35 mm spacers were used for 'T-tracking' (section 3.2.5.) and sequencing (section 3.2.6.). The smaller of the two plates was silanised by spreading one to two ml of silanising solution (2% dimethyldichlorosilane in trichloroethane) over the inner plate surface and allowed to dry at room temperature. After assembling the glass plates, approximately 50.0 ml of sequencing polyacrylamide (6% acrylamide, 8M urea in 1x TBE, 300 μl of 10% ammonium persulphate, 40 μl TEMED) were poured using a pipette. The gels were allowed to polymerise for one hour. Immediately before loading the wells were flushed out with a pasteur pipette to remove free urea and unpolymerised acrylamide. Two μl samples were loaded onto two 6% acrylamide, 8M urea gels using a drawn-out glass capillary. The gels were electrophoresed in 1x TBE, at 30 mA/1.2 kV until the xylene cyanol dye had migrated half way down the gel (1.5 hours for a 'short run' gel) or the whole length of the gel (4 hours for a 'long run' gel).

When electrophoresis was complete, the glass plates were gently prised apart leaving the gel on the unsilanised plate. This was immersed in 10% acetic acid, 10% ethanol for 10 to 15 minutes to fix the DNA bands and to remove the urea. The gel/plate were then drained and the gel was transferred to a sheet of Whatman 3mm paper. This was covered with 'Gladwrap' and dried under vacuum at 80^{o}C. The gel was used to expose an X-ray film ('Cronex 4', Du Pont) at room temperature overnight. The resulting autoradiograph was developed (Maniatis *et al.*, 1982) and the sequence read.

3.2.5. TEMPLATE SCREENING (BANKIER *et al.*, 1987)

'T-tracking' (ddTTP screening) allows the identification of clones from a shotgun preparation which contain the same insert sequence. A single termination reaction (eg. ddT) is performed for each clone. For screening 10 clones, 2 pmol of 17-nucleotide universal sequencing primer, 4 μl of 10x annealing buffer (100mM Tris-HCl pH 8.5, 50mM $MgCl_2$) and 30 μl of dH_2O were mixed in an eppendorf tube, and 3 μl of mix aliquoted onto the side of 10 eppendorf tubes. Two μl of template DNA were aliquoted to the rim of each tube (1 clone/tube) and the tubes spun briefly to mix. The reactions were boiled for 30 seconds and allowed to cool to room temperature to facilitate hybridisation of the template DNA and primer.

In a separate eppendorf, 10 μCi of [α-^{35}S]dATP (400 Ci mmol^{-1}), 20 μl of d/ddT termination mix (appendix A) and 5 units of Klenow DNA polymerase I were added. Two μl of the termination/Klenow mix were then aliquoted into each tube of template DNA/primer and mixed to begin the polymerase reaction. After 15 minutes incubation at 37^{o}C, 2 μl of 0.125mM dNTP chase (0.125mM dATP, dCTP, dGTP and dTTP) were added and the tube incubated for a further

30 minutes at 37^{o}C. This allowed unterminated nascent chains to be extended. Before loading, 5 μl of formamide solution (98% deionised formamide, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 20mM EDTA) were added to each reaction and boiled for 2.5 minutes to denature the DNA.

3.2.6. DIDEOXY SEQUENCE REACTIONS (SANGER *et al.*, 1977)

See Fig. 3.2. for a summary of the sequencing reactions. For each template to be sequenced, 0.5 pmol of universal sequencing primer, 1 μl of 10x annealing buffer and 7 μl of dH_2O, were added to an eppendorf tube. Five μl of single stranded M13 template DNA were added to the tube. The primer and template DNA were denatured by heating to 80^{o}C for three minutes and annealed by cooling slowly to room temperature, then placed on ice. One unit of Klenow polmerase was added to each annealed template/primer mix, and 2 μl of the template DNA/polymerase mix were aliquoted into each of four eppendorf tubes.

Ten μCi of [α-^{35}S]dATP (400 Ci mmol^{-1}) were added in each of four eppendorf tubes. Ten μl of d/ddA, d/ddG, d/ddC and d/ddT termination mixes (appendix A) were added (one mix per tube). Two μl of the C termination mix/α^{35}S-dATP were aliquoted into the first tube, keeping it separate from the DNA/polymerase aliquots. This was repeated with the T, A and G mixes being aliquoted to the second, third and fourth tubes respectively. Each tube was spun briefly to facilitate mixing and incubated at 37^{o}C for 15 minutes. Two μl of 0.125mM dNTP chase solution were added to each tube and incubated for a further 30 minutes at 37^{o}C. Samples were prepared for loading as in the template screening procedure.

3.2.7. COMPUTER ANALYSIS OF SEQUENCE DATA

Nucleic acid sequence data were assembled using VAX/VMS software written by Peter Stockwell (Department of Biochemistry, University of Otago, Dunedin, New Zealand) and with Genesys software written for an IBM/PC compatible computer developed by W. Bottomley (CSIRO, Division of Plant Industry, Canberra, Australia).

Computer analysis, including a codon usage profile, was done using these software packages, and the DNA and Protein Sequence Analysis Programmes, ver. 4.2., developed for an IBM/PC compatible computer by Mount and Conrad (1986). Hydropathy was analysed using a programme written by G. Timmerman (CRD, DSIR, Private Bag, Christchurch). Secondary structure predictions were according to Garnier (1978) using a procedure combining SAS and Lotus 123, developed by G. Timmerman. Phylogenies were inferred using the Phylogeny Inference Package (Phylip), ver. 3.0., written by J. Felsenstein (1987) using a protein parsimony programme (PROTPARS).

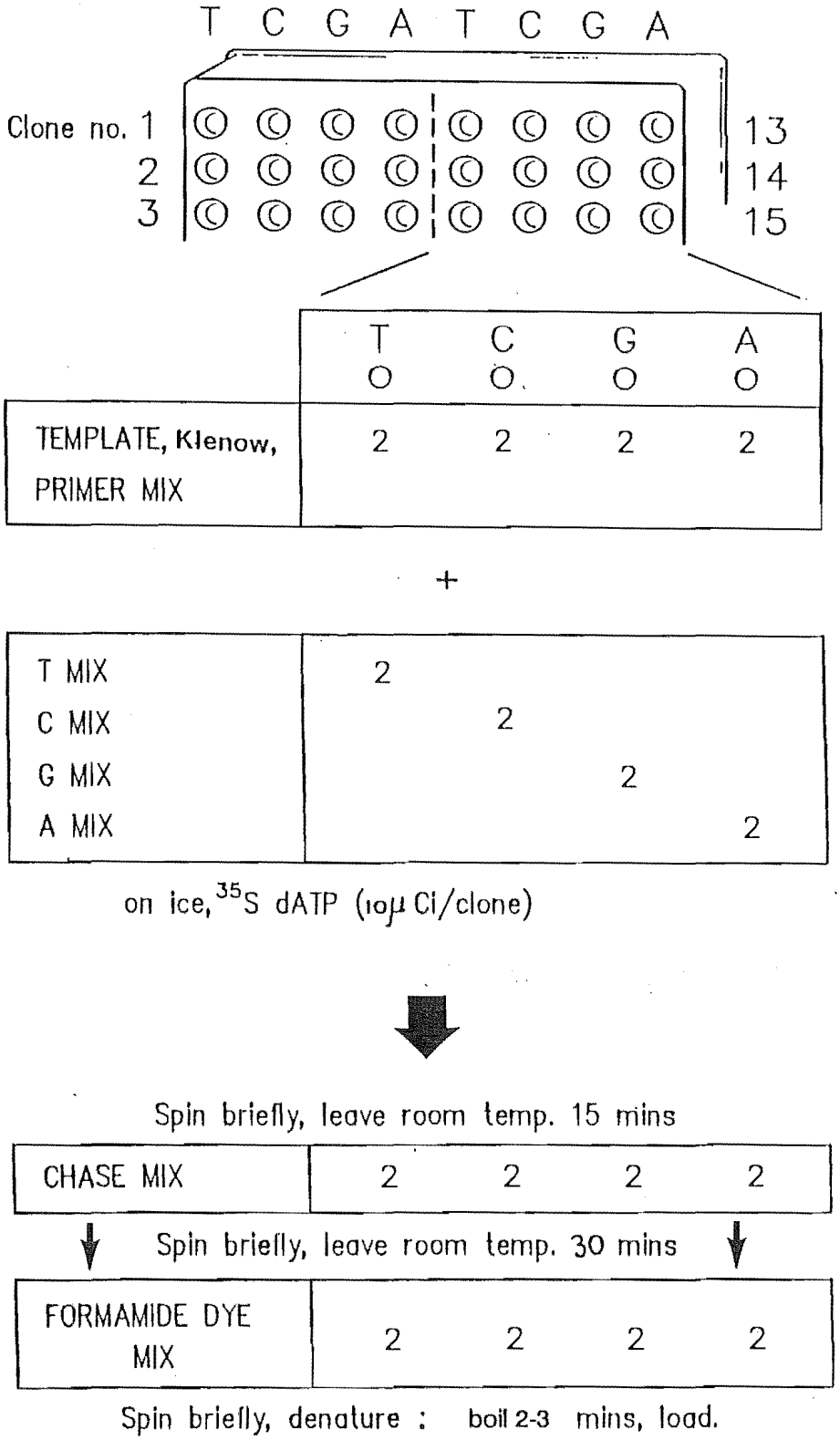


Figure 3.2. Summary of sequencing reactions (after Bankier *et al.*, 1987).

3.2.8. AMINO ACID SEQUENCING

The first ten amino acids of PVY^N coat protein were sequenced by Alan Carne (Department of Biochemistry, University of Otago, Dunedin, New Zealand), using a modified Edman degradation with a gas phase sequenator. The coat protein was purified (section 2.2.2.) and dialysed against dH_2O overnight to remove residual salt. One μg was analysed by SDS-PAGE and stained with coomassie blue and silver stain (section 2.2.8.) to confirm the purity of the preparation. Two hundred pmoles in 100 μl of dH_2O were sent for analysis.

3.3. RESULTS

3.3.1. FRAGMENT ISOLATION AND CLONING

Two methods were used to generate fragments of PVY^N insert cDNA to be sequenced. The first used restriction enzymes to cleave the whole plasmid, and insert DNA was separated from pUC19 plasmid bands on a polyacrylamide gel. The second method required the *Xba*I insert fragment of pVYN27 to be purified and sub-fragments were obtained from this by digesting the isolated fragment with a number of restriction endonucleases.

PVYN27 Insert

pVYN27 was digested with *Xba*I, *Hae*III, *Taq*I and *Rsa*I restriction endonucleases, the insert fragments isolated from a polyacrylamide gel and the DNA eluted. In this way, the fragments listed in Table 3.2. were purified and were cloned into M13 sequencing vector.

Table 3.2. Fragments generated by the digestion of PVYN27 with specific restriction endonucleases.

R.E.	Fragments[a]	Length (bp)[b]
*Xba*I	1	1200
*Hae*III	2	480, 650
*Taq*I	4	170, 190, 200, 270
*Rsa*I	3	150, 280, 430

a. Number of fragments purifed from the 6% polyacrylamide gel.

b. The approximate fragment length estimated from the DNA markers (BRL).

The ligated M13/insert fragments were transformed into the TG1 strain of *E. coli* with an overall high efficiency. Plate 3.1. shows the higher numbers of recombinant plaques (clear), compared with background non-recombinant plaques (blue) caused by re-circularisation of M13 vector. Control plates indicated a high efficiency of transformation into TG1.

The number of plaques picked for template preparation varied, depending on the fragmentation procedure. Where single fragments had been isolated and purified from an acrylamide gel (*Hae*III, *Rsa*I and *Taq*I), six plaques were routinely picked and single stranded template DNA prepared. Template DNA was purified from 20 recombinant plaques containing the *Xba*I fragment. Where sub-fragments of the insert had been cloned by 'shotgunning', 24, 47 and 23 templates were prepared from recombinant plaques containing *Alu*I, *Taq*I and *Hpa*II digested fragments respectively.

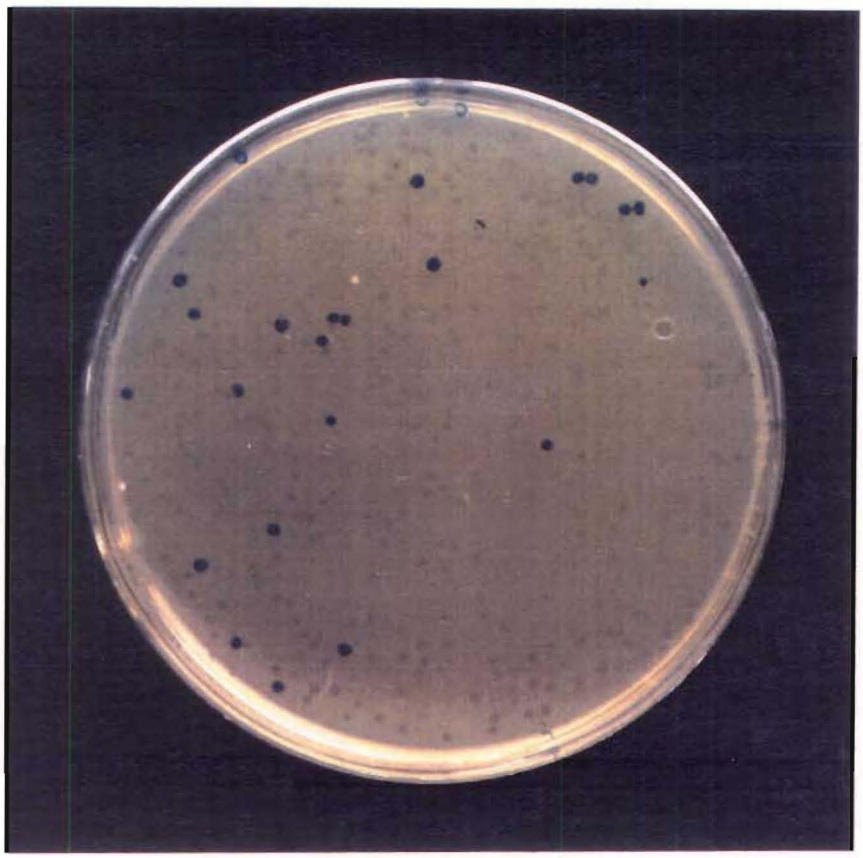


Plate 3.1. Transformation of TG1 (*E. coli*) with M13 containing pVYN27 inserts. Recombinant plaques are clear and non-recombinant plaques are blue

All the single stranded DNA isolated from recombinant phages gave good T-tracks and a number of these were sequenced further.

The overall strategy used to sequence the PVY^N insert is presented in Fig. 3.3. The sequence data obtained from each fragment and aligned to give an overall consensus sequence are presented in appendix B.

PVYC5 Insert

pVYC5 plasmid was digested with restriction endonuclease *Alu*I and six insert fragments were isolated from an acrylamide gel. The DNA eluted from the gel slices was cloned into the *Sma*I site in the dephosphorylated vector M13mp18. pVYC5 was also digested with restriction endonuclease *Xba*I and the fragment DNA ligated into the *Xba*I site on the M13 vector without further purification.

The recombinant phage vectors were transformed into *E. coli* strain TG1. Six plaques resulting from the transformation of recombinant phage containing each of the six *Alu*I fragments, and eight plaques from recombinant phage containing *Xba*I fragments of PVYC5 were selected for template preparation. These were T-tracked and a number of clones fully sequenced.

3.3.2. DIDEOXY SEQUENCING

PVYN27 Insert Sequence

The sequence of the 1134 nucleotide cDNA insert in plasmid pVYN27 was determined by the dideoxy chain termination method (Sanger *et al.*, 1977). Plate 3.2. shows an autoradiograph of a 'short' DNA sequencing gel and shows the junction between the vector DNA sequence and the 3'-terminal DNA sequence for the PVY^N coat protein gene.

Fig. 3.4. presents the 3'-terminal 1134 nucleotide sequence of PVY^N cDNA and the inferred amino acid sequence of 264 amino acids of the coat protein. Computer analysis revealed one large open reading frame (ORF) of 796 nucleotides in the plus-strand and an untranslated region 326 nucleotides long, followed by twelve nucleotides of polyadenylate. The remaining five reading frames contained numerous stop codons and no ORF of more than 60 codons on the plus-strand and 110 codons on the minus-strand.

The large ORF terminated at an opal stop codon (TGA), and was followed three nucleotides later by a second, in-frame, termination codon (TAG). This is capable of encoding a protein of 264 amino acids (Fig. 3.4.) with an estimated molecular weight of 29,631. The molecular weight is in reasonable agreement with the molecular weight (34 kd) for the PVY coat protein estimated by SDS-polyacrylamide gel electrophoresis (section 2.3.2.). No initiation codon (AUG) for the large ORF was identified, but by analogy with TEV (Allison *et al.*, 1986) and TVMV (Domier *et al.*, 1986), this is probably situated near the 5'-terminus of the genome and therefore is not part of the sequence being analysed. Comparisons with the amino acid

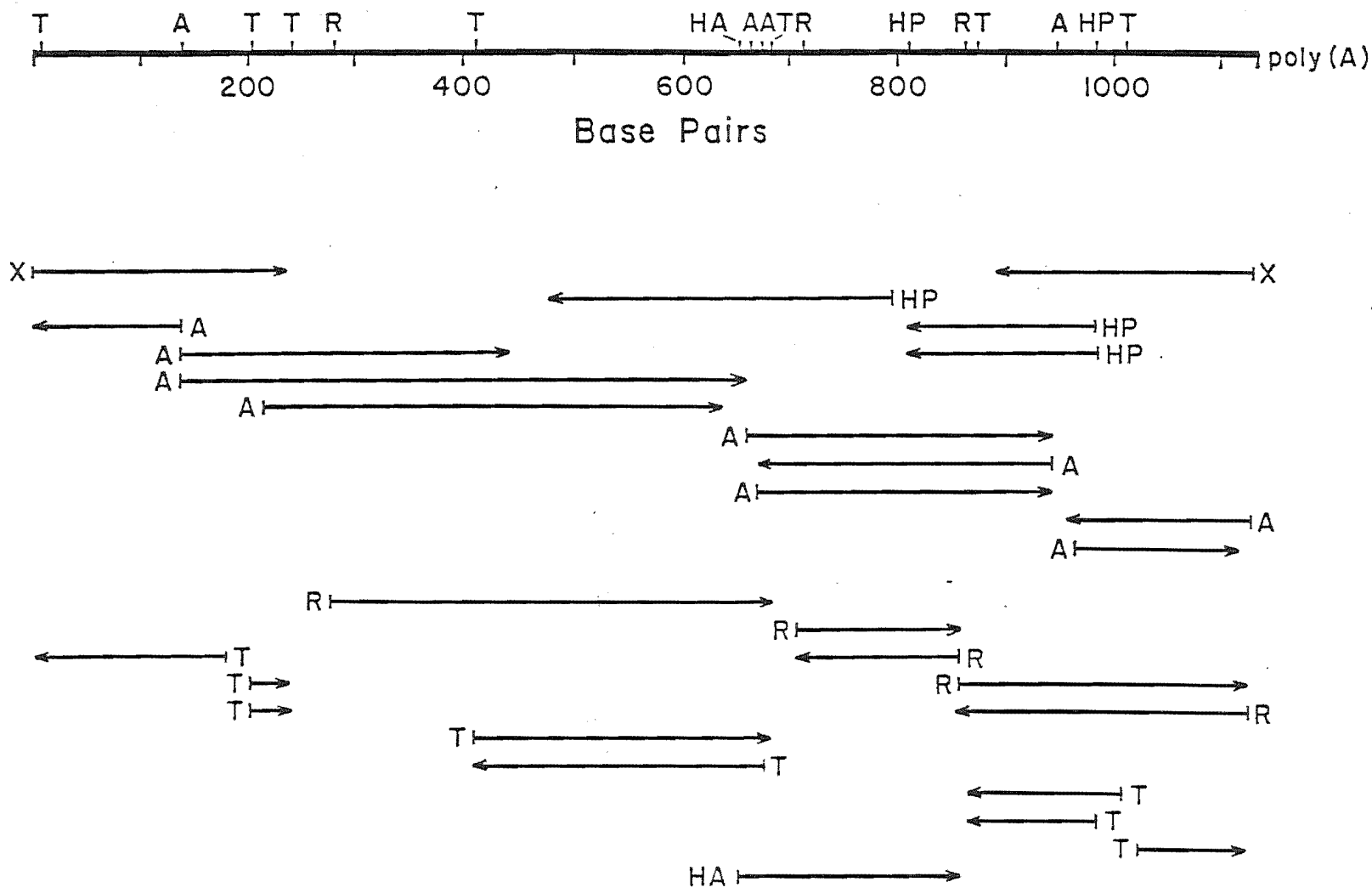


Figure 3.3. Strategy for sequencing 1134 nucleotides of the PVY^N coat protein gene. The direction and length of sequencing each fragment is indicated by the arrowed lines. Fragments were generated using the following restriction endonucleases: *Xba*I (X), *Taq*I (T), *Rsa*I (R), *Alu*I, *Hpa*II (HP), and *Hae*III (HA). A partial restriction map is shown at the top of the figure.

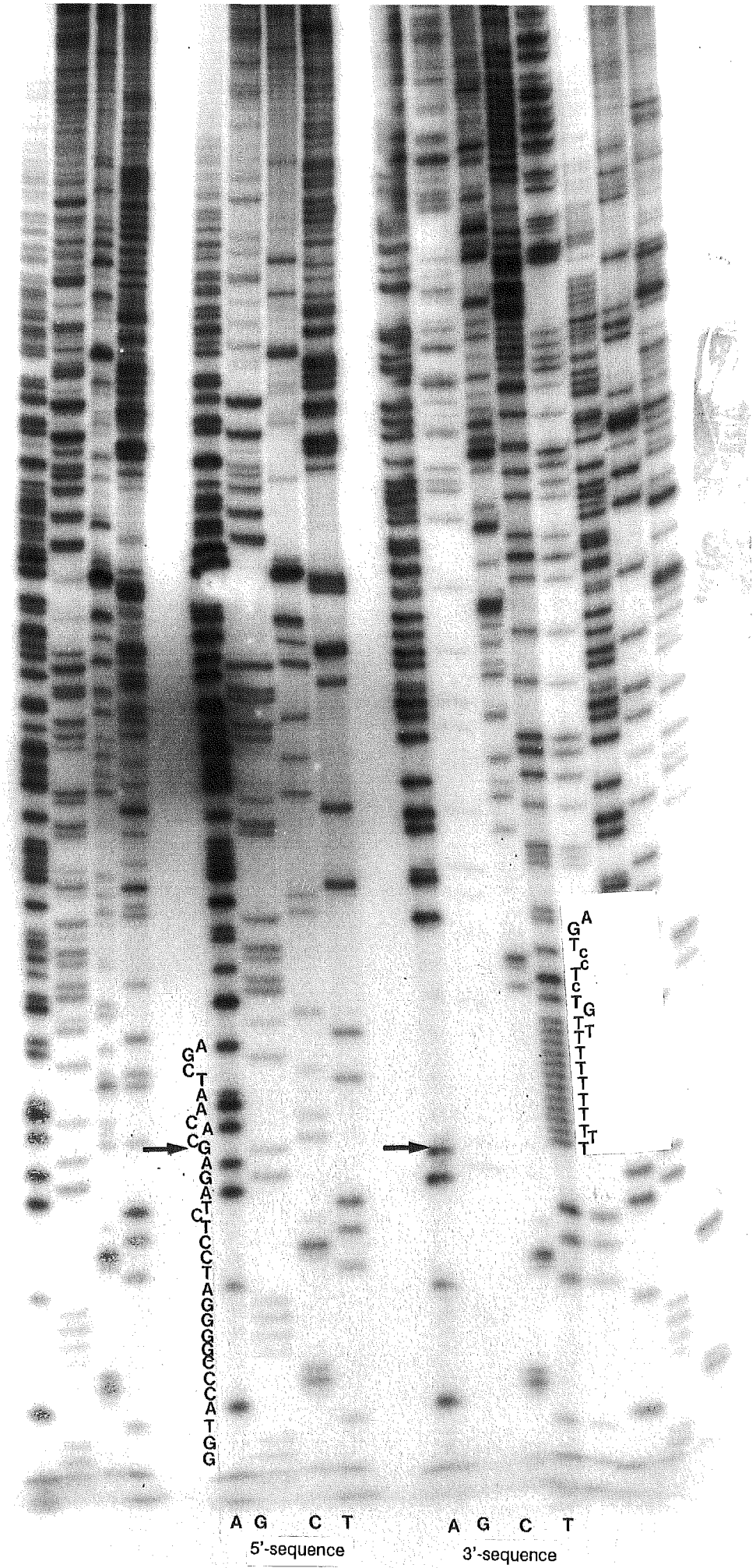


Plate 3.2. A 'short' sequencing gel. The junction between the vector DNA (*Xba*I- AGATCT) and the viral cDNA sequence is indicated.

```
1 C ACA ATC GAT GCA GGA GGA AGC ACT AAA AAG GAT GCA AAA CAA GAG GAA GGT
    T   I   D   A   G   G   S   T   K   K   D   A   K   Q   E   Q   G  17

53 AGC ATT CAA CCA AAT TTC AAC AAG GAA AAG GAA AAG GAC GTG AAT GTT GGA
    S   I   Q   P   N   F   N   K   E   K   E   K   D   V   N   V   G  34

104 ACA TCT GGA ACT CAT ACT GTG CCA CGA ATT AAA GCT ATC ACG TCC AAA ATG
     T   S   G   T   H   T   V   P   R   I   K   A   I   T   S   K   M  51

155 AGA ATG CCC AAG AGT AAA GGT GCA ATT GCA TTA AAT TTG GAA CAC TTA CTC
     R   M   P   K   S   K   G   A   I   A   L   N   L   E   H   L   L  68

206 GAG TAT GCT CCA CAG CAA ATT GAC ATC TCA AAT ACT CGA GCA ACT CAA TCA
     E   Y   A   P   Q   Q   I   D   I   S   N   T   R   A   T   Q   S  85

257 CAG TTT GAT ACG TGG TAT GAA GCA GTA CAA CTT GCA TAC GAC ATA GGA GAA
     Q   F   D   T   W   Y   E   A   V   Q   L   A   Y   D   I   G   E 102

308 ACT GAA ATG CCA ACT GTG ATG AAT GGG CTT ATG GTT TGG TGC ATT GAA AAT
     T   E   M   P   T   V   M   N   G   L   M   V   W   C   I   E   N 119

359 GGA ACC TCG CCA AAC ATC AAC GGA GTT TGG GTT ATG ATG GAT GGA GAT GAA
     G   T   S   P   N   I   N   G   V   W   V   M   M   D   G   D   E 136

410 CAA GTC GAA TAC CCA CTA AAA CCA ATC GTT GAG AAT GCA AAA CCA ACA CTT
     Q   V   E   Y   P   L   K   P   I   V   E   N   A   K   P   T   L 153

461 AGG CAA ATC ATG GCA CAT TTC TCA GAT GTT GCA GAA GCG TAT ATA GAA ATG
     R   Q   I   M   A   H   F   S   D   V   A   E   A   Y   I   E   M 170

512 CGC AAC AAA AAG GAA CCA TAT ATG CCA CGA TAT GGT TTA GTT CGT AAT CTG
     R   N   K   K   E   P   Y   M   P   R   Y   G   L   V   R   N   L 187

563 CGC GAT GGA AGT TTG GCT CGC TAT GCT TTT GAC TTT TAT GAA GTT ACA TCA
     R   D   G   S   L   A   R   Y   A   F   D   F   Y   E   V   T   S 204

614 CGG ACA CCA GTG AGG GCT AGA GAG GCA CAC ATT CAA ATG AAG GCC GCA GCT
     R   T   P   V   R   A   R   E   A   H   I   Q   M   K   A   A   A 221

665 TTA AAA TCA GCT CAA TCT CGA CTT TTC GGA TTG GAT GGT GGC ATT AGT ACA
     L   K   S   A   Q   S   R   L   F   G   L   D   G   G   I   S   T 238

716 CAA GAG GAA AAC ACA GAG AGG CAC ACC ACC GAG GAT GTT TCT CCA AGT ATG
     Q   E   E   N   T   E   R   H   T   T   E   D   V   S   P   S   M 255

767 CAT ACT CTA CTT GGA GTG AAG AAC ATG TGA TTGTAGTGTCTTTCCGGACGATATATA
     H   T   L   L   G   V   K   N   M   *                           264

824 GATATTTATGTTTGCAGTAAGTATTTTGGCTTTTCCTGTACTACTTTTATCGAAATTAATAATCGTT

891 TGAATATTACTGGCAGATAGGGGTGGTATAGCGATTCCGTCGTTGTAGTGACCTTAGCTGTCGTTTC

958 TGTATTATTATGTTTGTATAAAAGTGCCGGGTTGTTGTTGTTGTGGCTGATCTATCGATTAGTTGAT

1025 GTTGCGATTTGTCGTAGCAGTGACTATGTCTGGATTTAGTTAGTTGGGTGATGCTGTGATTCTGTCA

1092 TAGCAGTGACTGTAAACTTCAATCAGGAGACAAAAAAAAAAAA
```

Figure 3.4. Nucleotide sequence (shown as DNA) of the 1134 nucleotides of the 3'-region of PVY^N, with nucleotides numbered from the 5'-terminus. The predicted amino acid sequence of the large ORF is presented below the nucleotide sequence in the standard single-letter code. The stop codon is shown as '*'.

sequence of the coat protein of a related strain, PVY^D (Shukla *et al.*, 1986), and with the nucleotide sequence of PeMV (Dougherty *et al.*, 1985) and PVY^{N*} (van der Vlugt *et al.*, 1989) indicated that the 1134 nucleotide sequence is probably missing eight nucleotides encoding the amino terminal amino acids of the mature capsid protein. See appendix C for a list of restriction enzymes which recognise sites in the PVY^N coat protein sequence.

A codon usage profile for PVY^N coat protein is shown in Table 3.3. Of the 64 available codons, 58 are used to encode the coat protein. Only one codon for cysteine (TGC) and two codons for proline (CCC, CCA) are used. Poorly represented codons within the coat protein sequence are cysteine (0.4%), tryptophan (1.2%) and phenylalanine (2.3%). Alanine (8.2%) and glutamine (8.2%) are the most highly represented of the twenty amino acids. These data show a predominantly higher percentage adenine and thymidine (A+T) than guanine and cytosine (G+C) content in the third degenerate base, with 36.6%, 29.4%, 19.2% and 14.7% of the codons terminating ending in A, T, G and C respectively. This correlates with the overall base composition for the coding region. The least favoured codon is XCG (1.5%) and the avoidance of the CG dinucleotide is also observed for plants and other eukaryotes (Murray *et al.*, 1989). The doublet TA is also avoided in codon positions II and III in most eukaryotes, dicotyledonous and monocotyledonous plants (Murray *et al.*, 1989). It is avoided in the PVY^N coat protein sequence, as are the dinucleotides CC, GC, GG and GT.

Adjacent to the 3' terminal end of the coat protein coding region of PVY^N RNA is an untranslated region which is 326 nucleotides long and terminates in a poly(A) tract. The 3'-untranslated region has 19.1% adenine and 44.0% uracil, with few cytosine (12.4%), and guanine (24.2%) bases. By contrast, the base composition of the coding region has an adenine content of 35.2%, with fewer cytosine (19.3%), guanine (22.7%), and uracil (22.8%) bases.

A comparison between the PVY^N coat protein sequence and those of PeMV (Dougherty *et al.*, 1985), TEV (Allison *et al.*, 1985b), TVMV (Domier *et al.*, 1986), JGMV (Shukla *et al.*, 1987; Gough *et al.*, 1987), PPV (Ravelonandro *et al.*, 1988), and PVY^D (Shukla *et al.*, 1986) is presented in Fig. 3.5. Considerable sequence similarities are observed, with the amino termini representing the most variable regions. This is echoed by the hydropathy profiles of all seven potyviruses, which are all similar and indicate hydrophilic proteins (Fig. 3.6.). On the basis of amino acid alignment, the highest degree of similarity was between PVY^N, PVY^{N*}, PVY^D, and PeMV; with 98% similarity between the New Zealand and Dutch strains of PVY^N, 92% similarity between PVY^N and PeMV , and 91% between PVY^N and PVY^D, (Fig. 3.7.). The other potyviruses showed significantly less sequence similarity with PVY^N, for example PVY^N and TEV are only 62% similar. A pairwise percentage sequence similarity between eight potyviruses is presented in Fig. 3.8. An alignment of the nucleotide sequences of the 3'-termini of PVY^N and PeMV was made and this also showed substantial sequence similarities. There was an 88% sequence similarity in the capsid protein coding region (data not presented), and 83% in the 3'-untranslated region (Fig. 3.9.). Strains of PVY^N from New Zealand and the

Table 3.3. Codon usage profile for the coat protein of PVY^N coat protein.

TTT-Phe	3(1.1%)	TCT-Ser	3(1.1%)
TTC-Phe	3(1.1%)	TCC-Ser	1(0.4%)
TTA-Leu	4(1.5%)	TCA-Ser	5(1.9%)
TTG-Leu	3(1.1%)	TCG-Ser	1(0.4%)
CTT-Leu	5(1.9%)	CCT-Pro	0(0.0%)
CTC-Leu	1(0.4%)	CCC-Pro	1(0.4%)
CTA-Leu	2(0.8%)	CCA-Pro	12(4.5%)
CTG-Leu	1(0.4%)	CCG-Pro	0(0.0%)
ATT-Ile	7(2.6%)	ACT-Thr	8(3.0%)
ATC-Ile	6(2.3%)	ACC-Thr	3(1.1%)
ATA-Ile	2(0.8%)	ACA-Thr	7(2.6%)
ATG-MET	13(4.9%)	ACG-Thr	2(0.8%)
GTT-Val	9(3.4%)	GCT-Ala	7(2.6%)
GTC-Val	1(0.4%)	GCC-Ala	1(0.4%)
GTA-Val	1(0.4%)	GCA-Ala	12(4.5%)
GTG-Val	5(1.9%)	GCG-Ala	1(0.4%)
TAT-Tyr	7(2.6%)	TGT-Cys	0(0.0%)
TAC-Tyr	2(0.8%)	TGC-Cys	1(0.4%)
TAA-TER	0(0.0%)	TGA-TER	1(0.4%)
TAG-TER	0(0.0%)	TGG-Trp	3(1.1%)
CAT-His	3(1.1%)	CGT-Arg	1(0.4%)
CAC-His	3(1.1%)	CGC-Arg	3(1.1%)
CAA-Gln	11(4.2%)	CGA-Arg	4(1.5%)
CAG-Gln	2(0.8%)	CGG-Arg	1(0.4%)
AAT-Asn	8(3.0%)	AGT-Ser	4(1.5%)
AAC-Asn	6(2.3%)	AGC-Ser	2(0.8%)
AAA-Lys	9(3.4%)	AGA-Arg	2(0.8%)
AAG-Lys	8(3.0%)	AGG-Arg	3(1.1%)
GAT-Asp	9(3.4%)	GGT-Gly	4(1.5%)
GAC-Asp	4(1.5%)	GGC-Gly	1(0.4%)
GAA-Glu	14(5.3%)	GGA-Gly	11(4.2%)
GAG-Glu	7(2.6%)	GGG-Gly	1(0.4%)

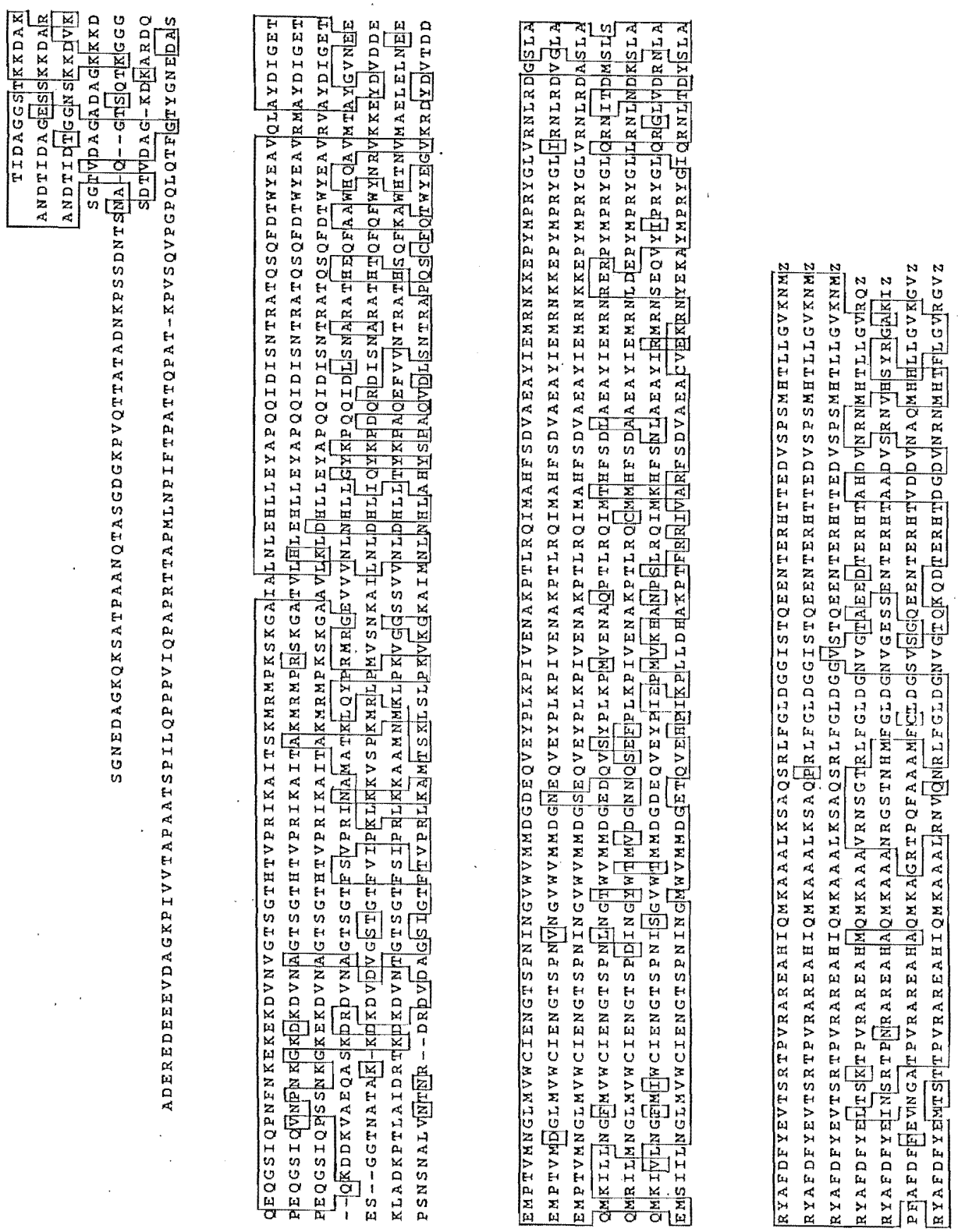


Figure 3.5. Alignment of the amino acid sequences of PVY^{N} (PVY*), PeMV, TEV, JGMV (SCMV), TVMV, PPV and PVY-D (PVY) coat proteins. The boxed residues and regions represent sequence homologies. The comparison was made on the basis of the predicted PVY^{N} coat protein data.

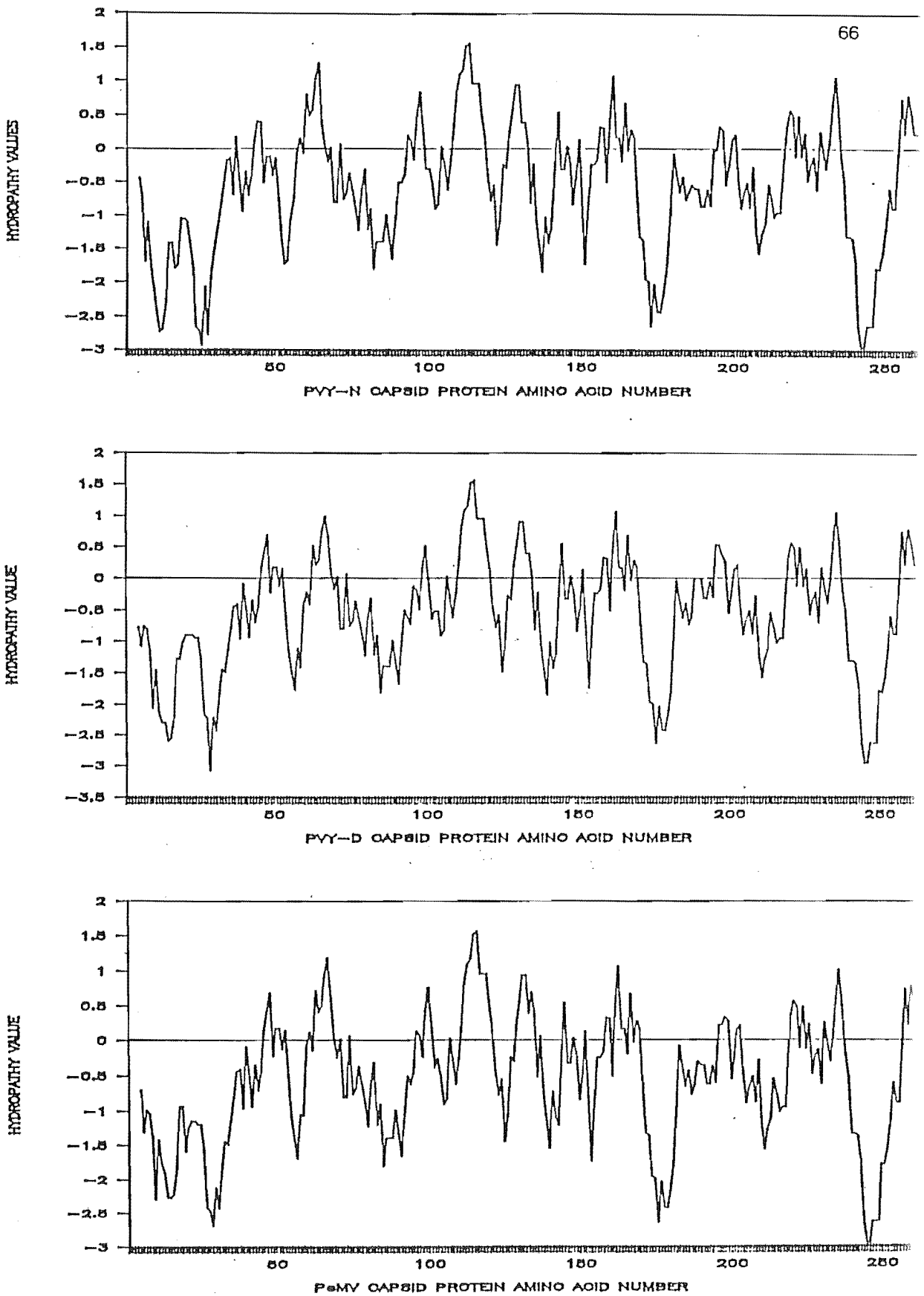


Figure 3.6. (A) A comparison of the hydropathy profiles of the coat protein sequences of PVY^{N}, PVY-D and PeMV. A sequence of nine amino acids was used to calculate the hydropathic value at each position as described by Kyte and Doolittle (1982). Scores are plotted from the N-terminus (position 1) to the C-terminus.

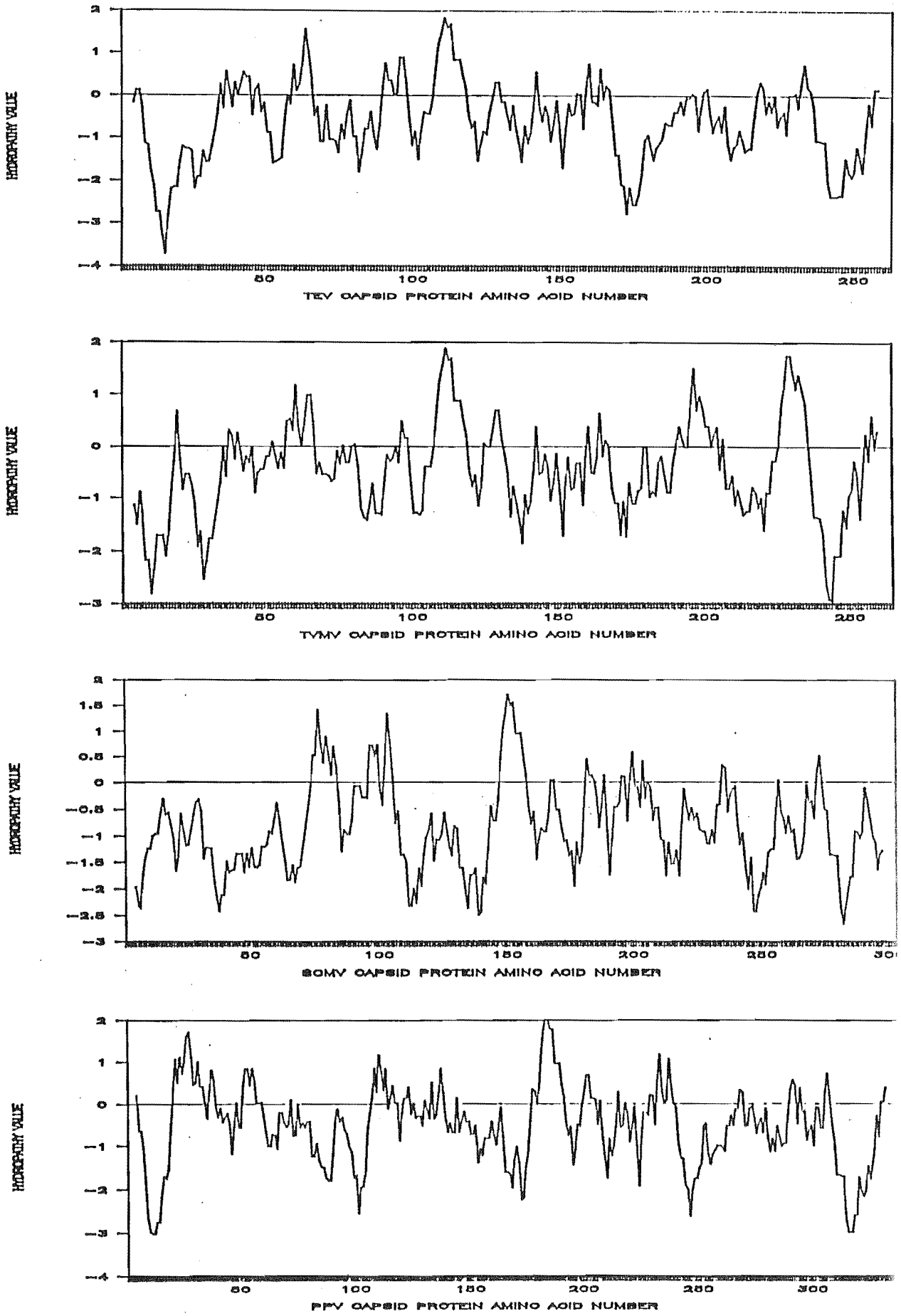


Figure 3.6. (B) A comparison of the hydropathy profiles of the coat protein sequences of TEV, TVMV, JGMV (SCMV) and PPV.

```
                10          20          30          40          50
PVY^N*  GNDTIDAGGS TKKDAKQEQG SIQPNLNKEK VKDVNVGTSG THTVPRIKAI
        *                     *           *
PeMV    ANDTIDTGGN SKKDVKPEQG SIQPSSNKGK EKDVNAGTSG THTVPRIKAI
        +     + + +   + +       ++  +       +
PVY^N   SNDTIDAGGS TKKDAKQEQG SIQPNFNKEK EKDVNVGTSG THTVPRIKAI
        #       #  #    ##       # #  #  #     #
PVY-D   ANDTIDAGES SKKDARPEQG SIQVNPNKGK DKDVNAGTSG THTVPRIKAI

                60          70          80          90          100
PVY^N*  TSKMRMPKSK GATVLNLEHL LEYAPQQIEI SNTRATQSQF DTWYEAVQLA
                   **                   *
PeMV    TAKMRMPKSK GAAVLKLDHL LEYAPQQIDI SNTRATQSQF DTWYEAVRVA
         +          ++ + +                                 ++
PVY^N   TSKMRMPKSK GAIALNLEHL LEYAPQQIDI SNTRATQSQF DTWYEAVQLA
         #     #    ## #                                    ##
PVY-D   TAKMRMPRSK GATVLHLEHL LEYAPQQIDI SNTRATQSQF DTWYEAVRMA

                110         120         130         140         150
PVY^N*  YDIGETEMPT VMNGLMVWCI ENGTSPNING VWVMMDGDEQ VEYPLKPIVE
PeMV    YDIGETEMPT VMNGLMVWCI ENGTSPNING VWVMMDGSEQ VEYPLKPIVE
                                                  +
PVY^N   YDIGETEMPT VMNGLMVWCI ENGTSPNING VWVMMDGDEQ VEYPLKPIVE
                   #                    #          #
PVY-D   YDIGETEMPT VMDGLMVWCI ENGTSPNVNG VWVMMDGNEQ VEYPLKPIVE

                160         170         180         190         200
PVY^N*  NAKPTLRQIM AHFSDVAEAY IEMRNKKEPY MPRYGLVRNL RDGSLARYAF
PeMV    NAKPTLRQIM AHFSDVAEAY IEMRNKKEPY MPRYGLVRNL RDASLARYAF
                                                     +
PVY^N   NAKPTLRQIM AHFSDVAEAY IEMRNKKEPY MPRYGLVRNL RDGSLARYAF
                                                  #     ##
PVY-D   NAKPTLRQIM AHFSDVAEAY IEMRNKKEPY MPRYGLIRNL RDVGLARYAF

                210         220         230         240         250
PVY^N*  DFYEVTSRTP VRAREAHIQM KAAALKSAQS RLFGLDGGIS TQEENTERHT
PeMV    DFYEVTSRTP VRAREAHIQM KAAALKSAQS RLFGLDGGVS TQEENTERHT
                                                  +
PVY^N   DFYEVTSRTP VRAREAHIQM KAAALKSAQS RLFGLDGGIS TQEENTERHT
                                          #
PVY-D   DFYEVTSRTP VRAREAHIQM KAAALKSAQP RLFGLDGGIS TQEENTERHT

                260
PVY^N*  TEDVSPSMHT LLGVKNMZ
PeMV    TEDVSPSMHT LLGVKNMZ
PVY^N   TEDVSPSMHT LLGVKNMZ
PVY-D   TEDVSPSMHT LLGVKNMZ
```

Figure 3.7. A comparison of the amino acid sequence homology between PVY^N, PVY^{N*} (Neth.), PeMV and PVY-D coat proteins. '*', '+' and '#' denotes variation between PVY^{N*}, PeMV and PVY-D, respectively, and PVY^N. The PVY^N sequence includes the three 5'-terminal amino acid sequences determined by amino acid sequencing. The percentage coat protein sequence homologies are: PVY^N/PeMV - 92.5%, PVY^N/PVY-D - 91% and PVY^N/PVY^{N*} - 97.8%.

	PVY^N	PVY^{N*}	PVY-D	PeMV	TEV	JGMV	TVMV	PPV
PVY^N	-	97	90.3	91.8	60.4	50.5	53.2	46.3
PVY^{N*}		-	91.4	92.2	61.2	50.3	55.1	47.5
PVY-D			-	91.8	60.3	49.3	53.2	46.2
PeMV				-	61.9	49.8	54.7	47.2
TEV					-	49.8	57.7	46.6
JGMV						-	46.1	44.6
TVMV							-	39.3
PPV								-

Figure 3.8. Percent amino acid homologies between the coat proteins of PVY^N, PVY^{N*} (Neth.), PVY-D, PeMV, TEV, JGMV, TVMV and PPV. The PVY^N sequence used is outlined in Fig. 3.4. and does not include the three amino-terminal amino acids. Sequence homologies greater than 90% are indicated in bold.

Netherlands showed 97.5% similarity in the 3'-untranslated region (data not shown).

Analysis of the nucleotide sequence of the 3'-untranslated region of PVY^N RNA revealed a number of features (Fig. 3.9.). These included three pairs of direct repeats of 8,9, and 11 nucleotides, and four regions of sequence similarity of 31 to 34 nucleotides. The extent of the similarity between the four regions varied from 53% to 78%. For comparison, the direct repeats and the three regions of sequence similarity found in the 3'-untranslated region of PeMV RNA (Dougherty *et al.*, 1985) are also included in Fig. 3.9.a. Three of the four regions of sequence similarity, are also observed for PVY^{N*} (van der Vlugt *et al.*, 1989) (data not presented). One directly repeated sequence 11 nucleotides long, and a third 8 nucleotide sub-sequence of the same repeat were also observed for PVY^{N*}. In addition, there were two possible stem and loop structures involving 3'-untranslated sequences of PVY^N RNA which can be predicted by applying the rules of Tinoco *et al.* (1973). Their structures are presented in Fig. 3.9.b. More stem-loop structures can be predicted to form within the capsid protein coding region of PVY^N RNA (data not presented).

Secondary structure predictions for PVY^N coat protein using Garnier's algorithm (1978) indicated nine distinct regions of α-helix, five sections of β-sheet and 14 regions of either coil or turn (Fig 3.10., Table 3.4.). Table 3.4. includes the secondary structure features for PVY predicted by Shukla and Ward (1988a) using Chou and Fasman's method (1974). Regions similarly assigned by both studies are in bold type for comparison. The coil and turn conformations have been presented together as Garnier describes 'coil' as a minor aperiodic conformation, and the study presented here found the assignment of a coil state to be frequently scattered within sequences afforded turn status.

Seven potyviral coat protein sequences were edited to remove the highly variable amino terminal sequences. The remaining aligned sequences were analysed using a protein comparison programme (Phylip manual, 1987) to obtain an evolutionary tree. Using a 'jumble' option the programme was run six times. During each run, the order in which the protein sequences were added was randomly chosen by the programme. Two most parsimonious trees were consistently produced and are illustrated in Fig. 3.11.

PVYC5 Insert Sequence

The sequences of nine cDNA fragments generated by *Alu*I digestion and one *Xba*I fragment of pVYC5 cDNA were determined by the dideoxy chain termination method. The relative positions of these fragments is presented in Fig. 3.12.

Comparisons with the nucleotide sequence for the whole genomic DNA of TVMV (Domier *et al.*, 1986) and TEV (Allison *et al.*, 1986) failed to find any sequence similarity. Sequence comparison with the coat protein sequence for PVY^N also failed to find any similarity and a number of short

A

```
PVY^N  UUGUAGUGUCUUU--CCGGACGAUAUAUAGAUAUUUAUGUUUUGUAGUAAGUAUUUU-GGC
PeMV   UUGUGCUGCCUCUCUCCGGACGAUAUAUAAGUAUUUUACAUAUGCAGUAAGUAUUUUUGGC

PVY^N  UUUUCCUGUACUACUUUUAUCGAAAUUAAUAAAUC-GUUUGAAUAUUACUGGCAGAUAGGG
PeMV   UUUUCCUGUACUACUUUUAUUGUAACUAGUAAAUCAGUUUGAAUAUUAUUCAUAAAUAGAG

PVY^N  GUGGUAUAGCGAUUCCGUCGUUGUAGUGACCUUAGCUGUCGUUUCUGUAUUAUUAUGUUU
PeMV   GUGGCAGGAUGAUUUCGUCAUUGUGGUGACUCUAUUUGUGAUUUCUGUAUUAUUAAGUUU

PVY^N  -GUAUAAAAGUGCCGGGUUGUUGUUGUUGUGGCUGAUCUAUCGAUUAGUUGAUGUUGCGA
PeMV   CAAAUAAAAGUGCCGGGUCGUUGUUGUUGUAGGUGAAAUAUCGAUUAGGUGAUGUUGCGA

PVY^N  UUU-GUCGUAGCAGUGACUAUGUCUGGAUUUAGUUAGUUGGGUGAUGCUGUGAUUCUGUC
PeMV   UUUUGUCGCAACAGUGACUAUGUCUGGUUCUACUUACUUGGGUGGUGUUAUGAUUUCGUC

PVY^N  AUAGGAGUGACUGUAAACUUCAAUCAGGAGACA poly(A)
PeMV   AUAACAGUGACUGUAAACUUCAAUCAGGAGACA poly(A)
```

B

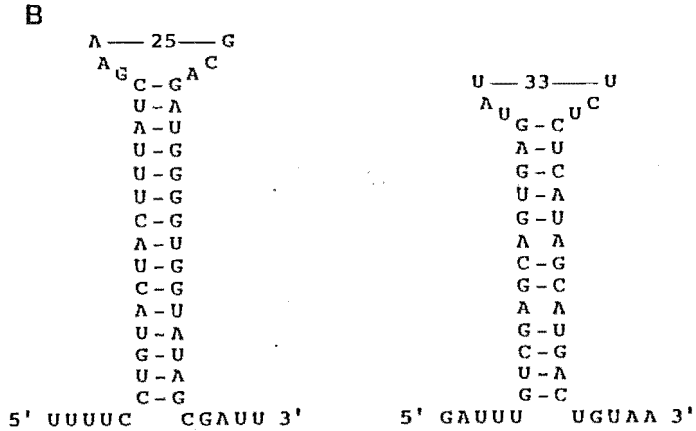


Figure 3.9. Analysis of the 3'-untranslated region of PVY^N RNA and alignment with PeMV RNA sequences. **(A)** Alignment between the 3'-terminal sequences of PVY^N and PeMV. Direct repeats occurring within these sequences are boxed, and regions of sequence similarity are underlined. **(B)** Secondary structures predicted for the 3'-terminal sequence of PVY^N RNA. Sequence 1 begins at nucleotide 854 and the free energy (ΔG) for the stem structure is predicted to be -14.2 kcal/mole. Sequence 2 begins at nucleotide 1030 and the predicted stem structure has a ΔG= -12.3 kcal/mole.

Table 3.4. Predicted secondary structure for the coat protein of Potato Virus Y. (A) PVY^N. The numbers refer to the amino acid residues shown in Fig. 3.10. (B) Secondary structure features predicted for PVY by Shukla and Ward (1988). Regions similarly assigned by both studies are in bold type in (A) and (B).

(A)

α-helix	β-sheet	Coil and Turn
15-19	43-49	5-11
29-34	**116-121**	12-14
50-57	**131-134**	20-28
60-73	**186-190**	35-42
99-108	202-206	74-94
135-142		109-115
157-177		122-130
196-201		143-147
214-231		151-156
		191-195
		207-213
		235-241
		245-250
		254-264

(B)

α-helix	β-sheet
46-50	**115-120**
62-74	**131-135**
74-80	**184-190**
87-93	257-261
95-100	
145-153	
155-163	
164-174	
209-229	
241-246	

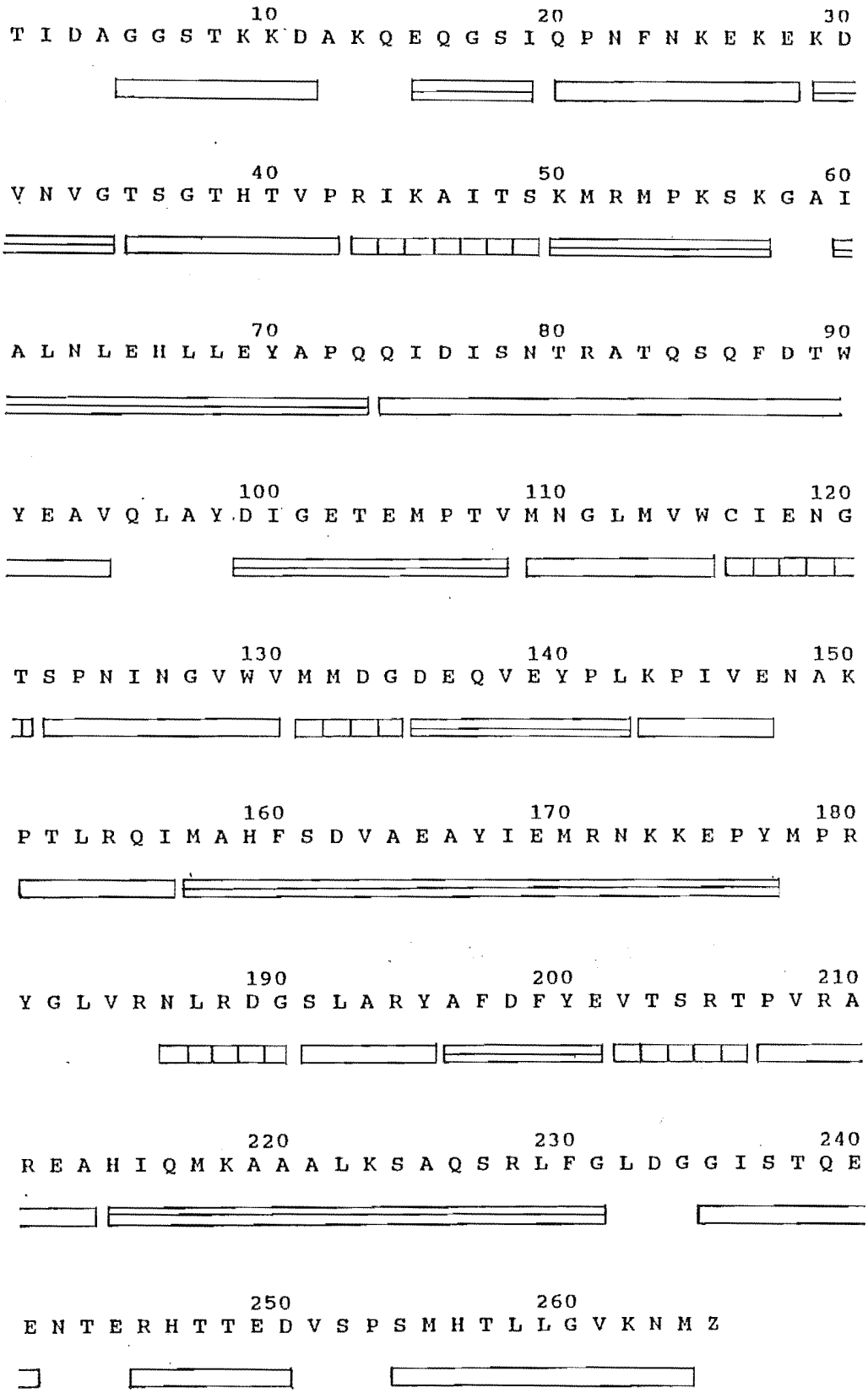


Figure 3.10. Secondary structure predictions for the PVY^N coat protein. Regions of α-helix, β-sheet and coil or turn are indicated by ▭, ▭▭▭ and ▭, respectively.

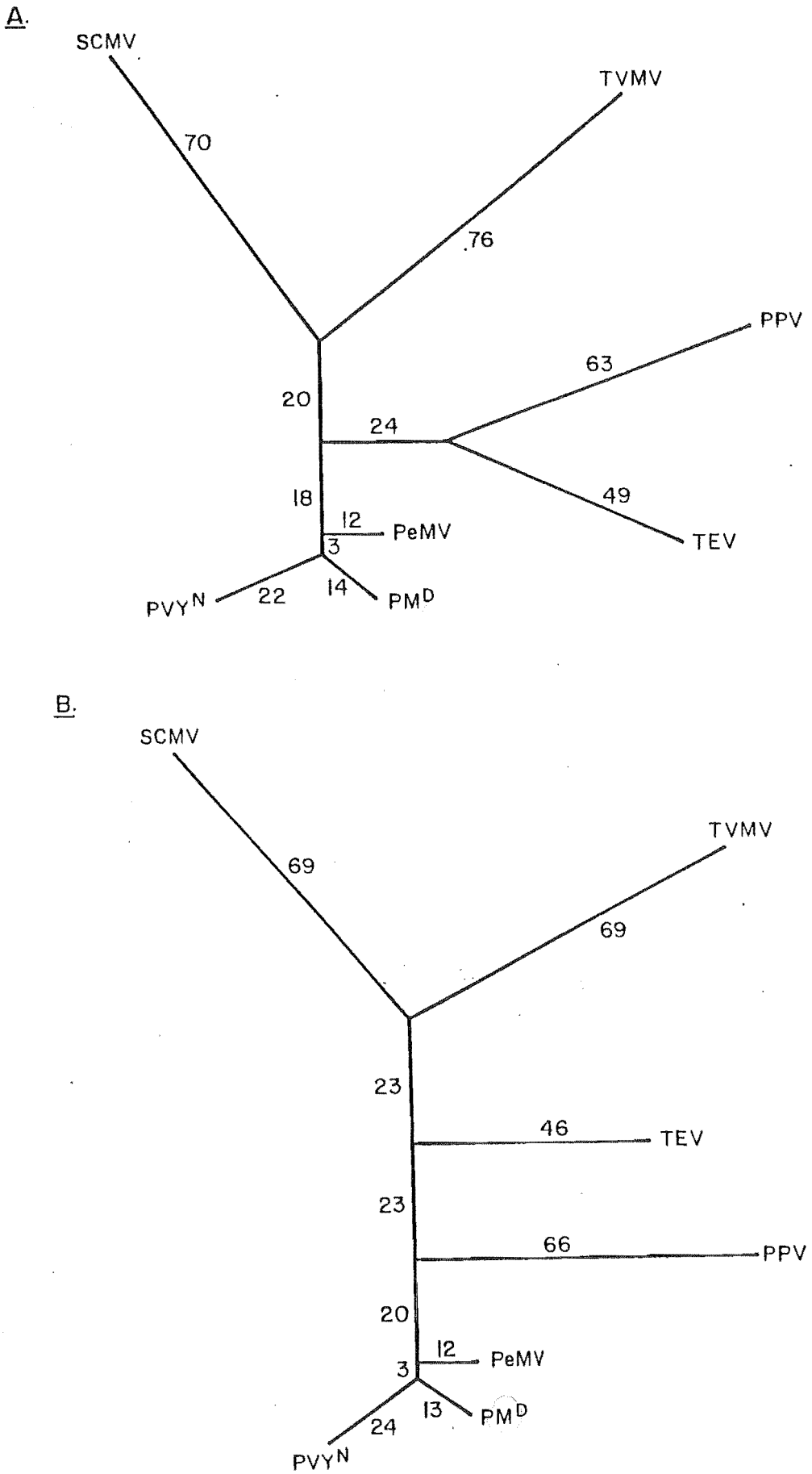


Figure 3.11. Phylogenetic analysis. The two most parsimonious trees are presented. The numbers of amino acid changes between each node are indicated.

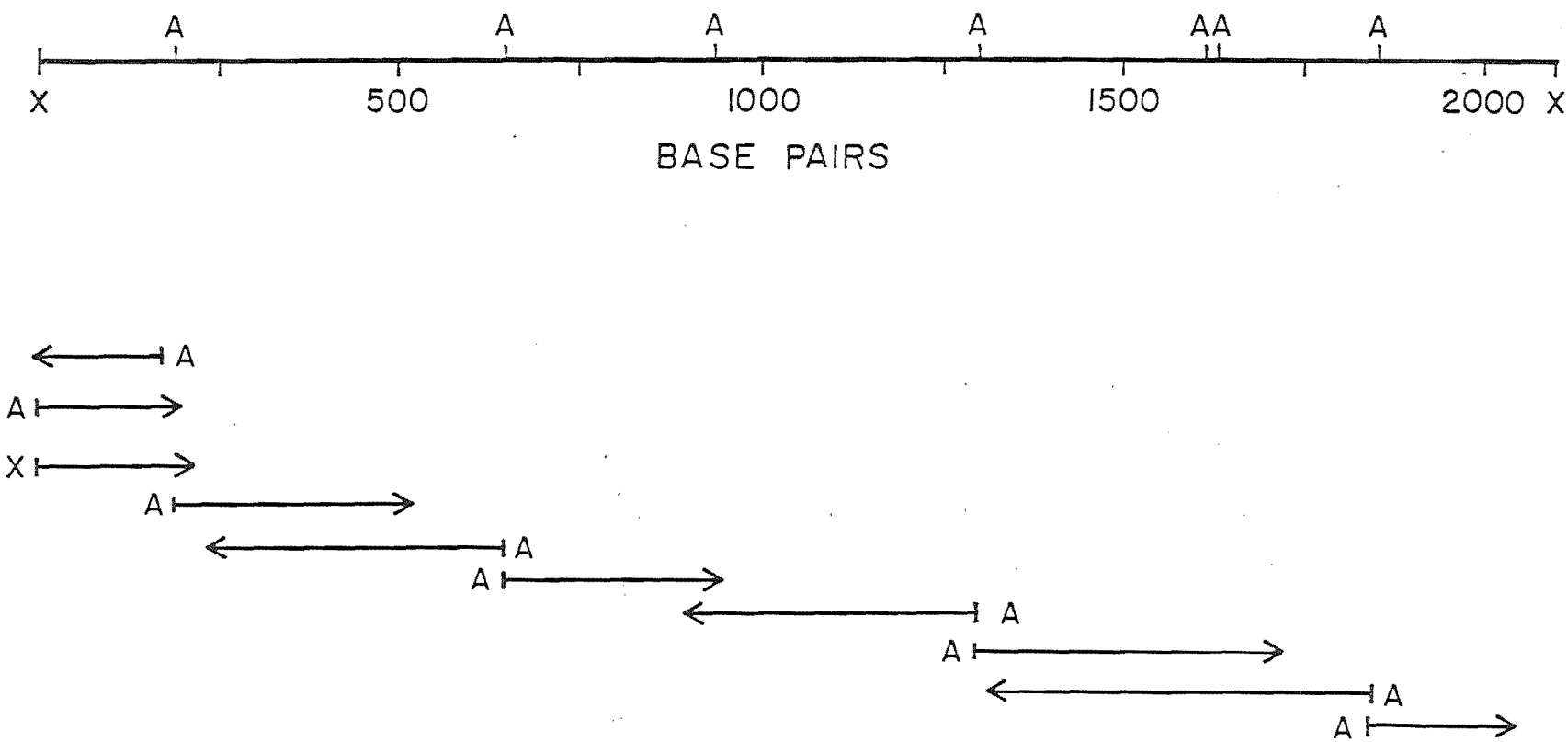


Figure 3.12. Strategy for sequencing selected fragments of the 2400 nucleotide PVY^{C} cDNA. The direction and length of sequencing each fragment is indicated by the arrowed lines. Fragments were generated using *Xba*I (X) and *Alu*I (A) restriction endonucleases. A partial restriction map is included at the top of the figure.

open reading frames were identified. This suggests that pVYC5 cDNA insert does not originate from anywhere in the potyviral genome and is possibly of host plant origin.

3.3.3. AMINO ACID SEQUENCING

PVY^N coat protein was purified from *N. tabacum* by the method of Reddick and Barnett (1983). It was confirmed as a pure preparation by SDS-PAGE and silver staining.

The first ten amino acids were sequenced by a gas phase sequenator and are presented in Fig. 3.13. The first amino acid serine was not confirmed and is therefore bracketed. The additional amino terminus information provided by amino acid sequencing overlaps and supports the amino acid sequence for the coat protein inferred from the DNA sequence data.

```
1       5       10
(S) N D T I D A G G S
```

Figure 3.13. The amino-terminal amino acid sequence of PVY^N coat protein

3.4. DISCUSSION

3.4.1. PVYN27

The nucleic acid and protein sequence information presented for the 1122 nucleotides of PVY^N indicated that the coat protein gene was adjacent to the 3'-terminus of the genome. This agreed with the published sequence data of other potyviral RNAs (Domier *et al.*, 1986; Ravelonandro *et al.*, 1988; Allison *et al.*, 1985b). The location of the carboxy terminal codons for the coat protein was predicted by comparison with the amino acid sequence of another strain, PVY^D (Shukla *et al.*, 1986), and by the presence of two in-frame termination codons. The location and identity of the amino terminal codon for PVY^N coat protein cannot be inferred from these sequence data, but amino acid sequencing identified the amino terminal residues. With the exception of the first amino acid residue, which could not be identified unambiguously (A. Carne, pers.comm), these were similar to the sequence published for PVY^D (Shukla *et al.*, 1986) and for PVY^{N*} (van der Vlugt *et al.*, 1989).

Potyviral coat proteins are known to be proteolytically cleaved from a longer precursor polyprotein at a glutamine/alanine (Q/A), glutamine/serine (Q/S) or glutamine/glycine (Q/G) dipeptide sequence. The amino acid sequence data for two other PVY strains, PVY^D (Shukla *et al.*, 1986) and a Dutch isolate PVY^{N*} (van der Vlugt *et al.*, 1989) suggest cleavage at Q/A and Q/G dipeptide sequences respectively. Protein sequence data for the New Zealand isolate of PVY^N tentatively indicates cleavage at the third dipeptide sequence option, Q/S. It therefore appears that strains of the same potyvirus may differ in proteolytic processing sites. Single site mutations in the genetic code could potentially account for this variation. The three nucleotide codon for glycine (GGA) could undergo a one nucleotide substitution in the second degenerate base to give rise to alanine, (GGA→GCA). A substitution in the first base of an alanine codon could result in serine (GCA→TCA). No nucleotide sequence data are available for PVY^D or PVY^N. However, the amino acid sequences of four Australian strains of PVY all have alanine as the first amino acid in the coat protein sequence (Shukla *et al.*, 1988d).

Because of the degenerate nature of the genetic code, only part of the variation contained in a gene is expressed in its protein (Murray *et al.*, 1989). Systematic codon preferences have been reported for bacteria, yeast, plant and mammalian genes (Murray *et al.*, 1989). In general, the PVY^N codon usage pattern most closely resembles that of unicellular organisms rather than man and other higher eukaryotes, due to it's overall preference for A+T content in codon position III. Chloroplast codon profiles also share this bias in the third degenerate base (Murray *et al.*, 1989). In unicellular organisms, highly expressed genes use a smaller subset of codons. However PVY^N tends to use almost all of the codons available, although some are more highly favoured than others. For example, 12 out of 13 proline residues are coded for by codon CCA.

Such 'codon bias' indicates an important factor to consider if modifying viral genes for high expression, for example within a plant system (see genetically engineered cross protection, section 4.1.3.).

The preference for A+T in codon position III is reflected by a high adenine content (35.2%) in the coding region. Previous reports suggest RNAs from potyviruses have a high adenylic content (Hill and Shepherd, 1972; Hill *et al.*, 1973; Knesek *et al.*, 1974) and this has been suggested to be a feature of potato-Y type viruses (Hill and Benner, 1976). Makkouk and Gumpf (1975), however, observed this not to be the case for several strains of PVY. This study observed a higher adenylate content compared to other bases for PVY^N and verifies the observation made by Hill and Benner (1976).

The seventeen potyviral coat proteins sequenced to date range in size from 263 amino acids (in TEV) to 330 amino acids (in PPV), mostly due to the difference in the length of their amino terminal peptide regions. Alignment of the amino acid sequences of six published potyvirus coat proteins shows extensive sequence similarity with the amino acid sequence predicted for the PVY^N coat protein. Some regions of these proteins contain blocks up to nine residues long that are identical in all seven capsid proteins. These regions are found in the middle and near the carboxy-termini of the polypeptides. In contrast, the amino terminal sequences were confirmed to be the least conserved in sequence and length. This may reflect different locations of specific cleavage sites in this highly variable region (Allison *et al.*, 1986; Domier *et al.*, 1986). For example, PPV capsid protein is 73 amino acids longer than PVY^N at the amino terminal end (Ravelonandro *et al.*, 1988), and JGMV is 42 residues longer (Shukla *et al.*, 1987). Only the two PVY^N isolates, PVY^D and PeMV showed a significant degree of homology in the amino terminal regions of their coat proteins.

High sequence similarities were observed between PeMV and PVY^N (92%), PVY^D and PeMV (92%), and between PVY^N and PVY^D (91%). Since this study was completed, the coat protein gene sequence of a Dutch strain of PVY^{N*} has been completed (van der Vlugt *et al.*, 1989). This strain showed 98% sequence similarity with the New Zealand isolate of PVY^N. Potyviruses have been classified using morphological and serological properties, however coat protein sequence similarities appear to be an increasingly useful criteria for resolving the complex taxonomy of the potyvirus group (Shukla and Ward, 1988b). The large percentage similarity between PVY^N, PVY^D and PeMV coat protein sequences suggested that PeMV should be considered a strain of PVY, and not a distinct member of the potyvirus group. This has also been suggested by Shukla *et al.* (1988d), who noted that PeMV showed high sequence similarity with four strains of PVY. Based on this, PeMV was assigned as a strain of PVY (Shukla *et al.*, 1986; Shukla and Ward, 1988b; van der Vlugt *et al.*, 1989). A lack of serological cross reaction between PeMV and PVY initially classified them as distinct species, and preliminary results suggest that this is due to a single sequence substitution in PeMV at a key residue in the major virus specific amino terminal epitope (Shukla and Ward, 1989).

The hydropathy profiles of the seven potyvirus coat proteins supported this re-classification. They showed that the PVY^N, PVY^D and PeMV hydropathy profiles were superimposable along their whole length. Differences in hydropathy were observed among the amino terminal residues of PPV, TEV, TVMV, JGMV and the PVY strains. Hydropathy profiles of the internal and carboxy-terminal sequences of all seven potyviruses examined had four strongly hydrophilic regions, and these were found in similar locations.

A preliminary alignment of the 3'-untranslated sequences of PVY^N and PPV, TEV, TVMV and JGMV showed no significant sequence similarities. The length of these regions varied considerably, from 475 nucleotides for JGMV (Shukla *et al.*, 1987) to 183 nucleotides for TEV (Allison *et al.*, 1986). However alignment of the 3'-untranslated nucleotide sequences of PVY^N and PeMV revealed much similarity. This observation supports the idea that analysis of the 3'-untranslated sequences may provide another method for distinguishing between strains and distinct members of the potyvirus group, and for determining their taxonomic relationships.

The major features of the 326 nucleotide long 3'-untranslated region of PVY^N RNA have been described (Fig. 3.9.) and these were compared with PeMV RNA and PVY^{N*} RNA. PVY^N has four regions of similarity in its untranslated sequence, and three of these four regions were found in the 3'-untranslated sequence of PeMV RNA and in PVY^{N*} RNA. Although both PVY^N and PeMV (Dougherty *et al.*, 1985) contain direct repeats, the location and sequences of these differed. An analysis of the possible secondary structural elements that may form in the 3'-untranslated region of PeMV RNA were also made. Two potentially stable stem structures could be predicted to form, however both have very long loop regions and neither correspond to the stem and loop structures that might potentially form in PVY^N RNA (data not presented). Possible base pairing structures have also been predicted by Dougherty *et al.* (1985) for two regions of the 3'-untranslated sequence of PeMV. Both have short stems, a three nucleotide loop, and relatively high Δ G values. A short stem-loop structure corresponding to one of the PeMV structures predicted by Dougherty *et al.* (1985) can be predicted in PVY^N with a Δ G = -6.8 for the stem. Although the location was the same, the sequence was different. Domier *et al.* (1986) predicted a similar stem-loop structure in the 3'-non-coding region of TVMV RNA but this was not in an equivalent position. No such structures were reported for TEV (Allison *et al.*, 1986). As a result of this analysis of PVY^N RNA, two potential regions of secondary structure were predicted in the 3'-untranslated region, but the potential to form similar structures was not found in the 3'-untranslated sequences of PeMV and TVMV. There were, therefore, no structural features such as direct repeats, repeated regions of sequence similarity, or possible secondary structures that were consistently observed in the 3'-untranslated regions of the six potyviruses whose nucleotide sequences are known. In another plus-strand RNA plant virus, brome mosaic virus (BMV), sequences displaying extensive secondary structure in the 3' end of the plus-strand have been shown to be involved in minus-strand synthesis (Dreher and Hall, 1988). Also in BMV, imperfect direct repeats have been implicated in promotion of subgenomic RNA synthesis (French and

Alquist, 1988). However, the biological significance of RNA secondary structure and direct repeats in potyviruses remains unclear.

The coat protein secondary structure predictions made in this study correlate with those made by Shukla and Ward (1988a). Garnier (1978) suggests that hydrophobic residues tend to cluster on the surface of an α-helix. For example, α-helices are observed for residues 50-57, 60-73, 99-108, 135-142 and 214-231, and all these regions are hydrophobic as indicated by the PVY^N hydropathy profile (Fig. 3.6.). Regions observed to be strongly hydrophilic by the hydropathy profile are assigned as either coil or turn, for example 20-28, 74-94, 143-147 and 245-250. Residues containing hydrophilic side chains include aspartic and glutamic acids (D and E), lysine (K), arginine (R), histidine (H), asparagine (N) and glutamine (Q). These are present in the hydrophilic sequences, 74-94: PQQIDISNTRATQSQFDTWY and 20-28: GSIQPNFNK. Chou and Fasman (1974) suggest that charged residues are unfavourable in β-sheet structures. Non-polar side chain contacts are formed with the hydrophobic α-helices, for example alanine (A), valine (V), leucine (L), isoleucine (I), phenylalanine (F), tryptophan (W) and methionine (M). These residues are highly represented in the α-helices 60-73, 99-108 and 214-232: GAIALNLEHLLEYA (60-73), AYDIGETEMP (99-108) and IQMKAAALKSAQ (214-232). Chou and Fasman cite leucine as the most abundant residue in the helical core and confer glutamaic acid (E), alanine (A) and leucine (L) as strong helix 'formers'. Glutamic acid was observed in seven out of the nine predicted α-helices and alanine residues were observed in the other two. Proline (P), asparagine (N) and glycine (G) are reported to be strong α-helix 'breakers' (Chou and Fasman, 1974). Residues commonly observed to form β-sheet conformations include valine (V), isoleucine (I) and tyrosine (T). All five sequences of amino acid residues assigned as β-sheet contained at least one valine residue.

Two unrooted, equally parsimonious phylogenetic trees were consistently produced using the modified coat protein data from seven potyviruses using the PROTPARS programme. The programme uses a method intermediate between the approaches of Eck and Dayhoff (1966) and Fitch (1971) and insists that any changes of amino acid be consistent with the genetic code (Felsenstein, 1987). The two trees differ in the positioning of TEV and PPV. As no outgroup could be defined and the trees are unrooted, no cladistic analysis could be undertaken. One other attempt to establish hierarchical relationships between distinct members of the potyvirus group was made by A. J. Gibbs *et al.* (Shukla and Ward, 1989). A dendrogram calculated from the coat protein sequence homologies of 25 strains of potyviruses was presented. This dendrogram, however, does not support either of the phylogenies produced in this study as being more or less likely.

3.4.2. PVYC5

The initial screening of recombinant plasmids was made by DNA slot blot hybridisation using PVY^N coat protein as a probe (section 2.3.6.). A number of fragments of pVYC5 were sequenced

and compared with the whole genomic sequences of TEV (Allison *et al.*, 1986) and TVMV (Domier *et al.*, 1986) to test whether the insert derived from viral genomic DNA other than the coat protein. No regions of nucleotide or amino acid sequence similarity were observed between the PVY^{C} fragments and the potyviral genomes. Sequence analysis of the inserted sequences, therefore confirmed that the plasmid pVYC5 did not contain sequences derived from PVY. It was concluded that the recombinant plasmid contained host plant sequences.

CHAPTER FOUR

PLANT TRANSFORMATION AND THE MOLECULAR ANALYSIS OF TRANSFORMANTS

4.1. INTRODUCTION

4.1.1. METHODS FOR CONTROLLING POTYVIRUSES

Potyviruses flourish in a wide range of host plants and environmental conditions. They present formidable disease problems in commercially and economically important crops, and currently, no effective antiviral agents exist for controlling plant viruses. In many countries the destruction of whole crops has been the only solution to virus infections (Edwardson, 1974). The use of both virus-free seeds and the isolation of growing areas from virus sources were advocated by Klinkowski and Schmelzer (1960) and Edwardson (1974) as important methods for maintaining healthy varieties of potato free from virus infection. However, where the virus may be transmitted to the progeny by seed, for example pea seed-borne mosaic virus, early sources of inoculum may be provided in the field. The isolation of infected areas is often compromised by weed-host reservoirs which may harbour plant viruses for long periods of time. Attempts at reducing the spread of a potyviral infection by controlling the vectors which transmit the viruses via insecticides and nematicides have also been largely unsuccessful (de Bokx and Huttinga, 1981; Edwardson, 1974). Host plants which are immune, resistant or hypersensitive to viruses may provide a basis for breeding programmes, but they may also aid in establishing more virulent virus strains (Edwardson, 1974). Another preventive measure is sterilising plant seeds, usually by heat (Valle *et al.*, 1988).

Edwardson (1974) outlines a number of schemes used internationally for maintaining PVY-free potato crops, and most are based on certified seed potato schemes. This is the current method of controlling PVY and other virus infections in New Zealand potato crops. Growers are required to maintain a virus-free line for three years before becoming eligible to provide seed for commercial distribution. This is achieved by 'roguing' or eliminating plants which show symptoms. The remaining plants are either naturally resistant, or are infected by mild strains of the virus and thus are 'cross protected'.

4.1.2. CROSS PROTECTION

Classical cross protection is a phenomenon whereby a plant pre-incubated with a mild virus strain (inducer) becomes resistant to subsequent inoculation by a related strain (challenger). The phenomenon was first reported between two strains of tobacco mosaic virus (TMV) by M^{C}Kinney in 1929. Similar observations were independently made by Thung in 1931 between TMV strains

(in Valle *et al.*, 1988), and by Salaman (1933) between strains of potato virus X (PVX) and later PVY. They all observed that the simultaneous inoculation of mild and severe virus strains in a host led to the suppression of severe viral symptoms.

The following characteristics have been attributed to the phenomenon. Cross protection is highly specific and is only observed among closely related viruses. Rare cases also exist where serologically related viruses have failed to induce cross protection (Valle *et al.*, 1988). Protection is often reciprocal and a given virus may be the inducer or challenger depending on the host (Zinnen and Fulton, 1986). The degree of protection appears to depend on the relative concentrations of the inducer and challenger. The host plants may also have a role in determining whether or not cross protection will ensue. Generally an increase in the concentration of a challenger results in the earlier expression of symptoms (Cassels and Herrick, 1977). Cross protection appears to be a relative, rather than absolute, phenomenon as the challenge virus may or may not be detectable in the protected plant (Cassels and Herrick, 1977; Dodds, 1982).

Cross protection has been proposed as a practical means of protecting plants against severe virus infection (Salaman, 1933). It was first used on a commercial scale in the Netherlands against TMV in tomato plants (Rast, 1972). Other examples of commercially applied cross protection are tomato mosaic virus in tomato and pepper in China, citrus tristeza virus in citrus crops in Brazil and Australia (Fulton, 1986) and papaya ringspot virus in papaya in Taiwan (Yeh *et al.*, 1988). In a local example, a mild strain of apple mosaic virus (AMV) isolated from the field, successfully cross protected apple orchards in New Zealand against the severe strains of AMV (Chamberlin *et al.*, 1964).

There are a number of potential disadvantages to using classical cross protection as a method of biological control (Valle *et al.*, 1988). It is both tedious and costly to perform. In addition, the protecting strain may spread to other sensitive species, or mutate to a more virulent strain (Valle *et al.*, 1988). It could potentially interact with other viruses in plants and synergistically lead to more severe symptoms. For example, infection by TMV or CMV individually causes a less severe disease in host plants than when they are present together (Palukaitis and Zaitlin, 1984). A classical case of synergism in potato occurs between PVY and PVX (Fulton, 1986).

There is no general consensus as to the possible physiological and molecular basis for classical cross protection and this is partly due to the diversity of virus-host systems that have been studied to date. The following models have been suggested and are reviewed by Valle *et al.* (1988), Hamilton (1980), Fulton (1980) and Sequeira, (1984). The 'antibody theory' proposed that virus infection led to the production of host substances which inhibited virus multiplication. Such protective compounds have been observed in virus infected plants but their appearance can not be correlated with cross protection (Sequeira, 1984). The 'precursor-exhaustion' theory proposes that cellular compounds such as metabolites, ribosomes and precursors for RNA and

protein biosynthesis are sequestered and used by the inducing virus, thus preventing multiplication of the challenging virus (Valle *et al.*, 1988). Other studies have implicated the viral coat protein in cross protection. The inducing strain (or free coat protein) may compete with the challenge virus for binding sites on the cell (Sherwood and Fulton, 1983), or the inducer may inhibit the uncoating of the challenge virus (Sherwood and Fulton, 1982). This hypothesis has not been tested in a number of virus-host systems; cross protection is observed among viroids and as they are non-capsidated, models based on coat protein can not apply. De Zoeten and Fulton (1975) suggested that the coat protein of the inducing virus may encapsidate the RNA of the challenger and thus prevent expression. Based on experiments which suggest that related viruses can use the same replicase, Gibbs (1969) and Ross (1974) proposed that the recognition of RNA by the replicase could be the basis of cross protection specificity. The replicase could specifically inhibit the expression of related viruses but not of unrelated ones. Finally, Palukaitis and Zaitlin (1984) suggested that the protective effect of cross protection is at the nucleic acid level. RNA complementary to the genomic RNA of the inducer may bind to the incoming RNA of the related challenger and thereby hinder expression.

No single model adequately accounts for cross protection and it must be assumed that the mechanism either varies depending on the virus-host system, or operates on a step in plant virus infection which has not yet been characterised. Possibly a number of mechanisms operate concurrently.

4.1.3. 'GENETICALLY ENGINEERED CROSS PROTECTION'

The advent of genetic engineering technology has provided new methods for obtaining plants resistant to disease. The transfer of genes into plant nuclear genomes can be achieved using the Ti plasmid of *Agrobacterium tumefaciens* (section 4.1.4.) and has proved successful for plant transformations (Herrera-Estrella *et al.*, 1984; Horsch *et al.*, 1985; Hoekema *et al.*, 1983; Bevan, 1984). DNA complementary to specific viral sequences has been integrated into appropriate hosts and used to manipulate and improve the virus resistance properties of various crops. Using this strategy, virus infection in plants has been controlled in two ways. One has been by integrating satellite RNAs into appropriate hosts (Baulcombe *et al.*, 1986). Baulcombe *et al.* (1986) described the transformation of tobacco with cDNA copies of CMV satellite RNA and the production of satellite RNA transcripts by these plants. Experiments showed symptoms in tobacco were markedly reduced when the inoculum contained satellite RNA acquired from transformed plants. Baulcombe (1989) suggested that satellite RNA in CMV infected plants competed with genomic RNA for a limited amount of replicase. A second method for controlling virus infection has been by the integration of viral coat protein genes into plants. This appears to mimic classical cross protection and in experimental situations has proved successful in inhibiting virus infection (Powell Abel *et al.*, 1986).

The coat protein gene of TMV was first manipulated into tobacco by Bevan *et al.* (1985) and was shown to be expressed at low levels. This has since been repeated, for example by Yamaya *et al.* (1988). To test the hypothesis that the free coat protein produced by an integrated TMV coat protein gene could inhibit TMV infection in a transgenic plant, Powell Abel *et al.* (1986) transformed the TMV coat protein gene into tobacco plants via an *Agrobacterium* vector. Transformed plants expressing high levels of coat protein were all found to delay expression of symptoms when challenged with either a mild or severe strain of TMV (Powell Abel *et al.*, 1986; Nelson *et al.*, 1987). The delay in disease development correlated with a lack of virus accumulation. Some plants never developed the disease, demonstrating that plants could be genetically transformed for resistance to virus disease development. Progeny of the transgenic plants inherited the coat protein gene in a Mendelian 3:1 segregation ratio (Powell Abel *et al.*, 1986). Nelson *et al.* (1987) observed similar results to classical cross protection studies performed by Sherwood and Fulton (1982) and Dodds (1985), both of whom used untransformed plants, in that protection was greater when whole virus rather than viral RNA was used as challenge inoculum. In addition, an increase in the concentration of the challenge virus inoculum correlated with a reduction in the time for symptom development. Therefore, expression of the coat protein gene seemed to mimic classical cross protection and supported a mechanism in which free coat protein provides protection. In contrast, plants demonstrating coat protein mediated cross protection for PVX (Hemenway *et al.*, 1988) and for TMV (Yamaya *et al.*, 1988) showed resistance to infection by viral RNA. This suggested a different mechanism for protection.

The following series of experiments implied that the virus coat protein and not the mRNA is responsible for protection against alfalfa mosaic virus (AlMV) infection. The coat protein cistron of AlMV and of tobacco rattle virus (TRV) were transformed into tobacco by van Dun *et al.* (1987). Translation of the AlMV coat protein could be detected in 70% of the transgenic plants. Plants accumulating AlMV coat protein were highly resistant to infection by a mixture of AlMV sub-genomic RNAs, whether or not the sub-genomic coat protein RNA was present (van Dun *et al.*, 1987). Similar observations were made for AlMV by Turner *et al.* (1987) and Loesch-Fries *et al.* (1987). Two non-structural genes (encoded on sub-genomic RNAs) involved in the replication of AlMV were individually transferred to the tobacco genome (van Dun *et al.*, 1988b). However no resistance to AlMV was observed for transgenics containing either of these proteins. In another experiment, an abbreviated mutant coat protein gene was transformed into tobacco (van Dun et al., 1988a). Although transcription of the gene was not affected, accumulation of the coat protein was prevented and resistance to infection with AlMV was abolished (van Dun *et al.*, 1988a).

Van Dun and Bol (1988) observed that a high degree of coat protein sequence homology was required for cross protection. Approximately 100% coat protein sequence homology occurs between a strain of TRV and pea early browning virus (PEBV), and cross protection was

conferred on transgenic TRV plants against PEBV. However, two strains of TRV which had 39% coat protein sequence similarity failed to show protection.

The antisense theory for classical cross protection has also been tested for genetically engineered cross protection. Transgenic tobacco plants containing cDNA encoding the PVX coat protein gene inserted in an antisense orientation, were found not to produce PVX coat protein (Hemenway *et al.*, 1988). Transgenic plants expressing PVX coat protein antisense transcripts were protected against infection by the virus, although at very low inoculum levels only.

There are a number of potential advantages to genetically engineered cross protection as opposed to classical cross protection. In genetically engineered protection, the inheritance of the gene is Mendelian (Horsch *et al.*, 1984) and the gene is permanently transferred to the progeny. This is not possible with classical cross protection. The level of protection afforded by genetic engineering could possibly be manipulated by increasing expression from multiple copies of the gene inserted in tandem into a host (Valle *et al.*, 1988). The high specificity of classical cross protection could be circumvented by inserting copies from more than one virus into a plant host (Valle *et al.*, 1988). Also, the risk of infection by inducing virus strains during the lag time is eliminated in genetically engineered protection (Valle *et al.*, 1988).

It is not known whether the mechanisms for classical and genetically engineered cross protection are the same. As in classical cross protection, free coat protein is strongly implicated as having a role in genetically engineered cross protection. Van Dun *et al.* (1988a) suggest there are three functional domains for AlMV coat protein. One of these domains may have a specific role relating to the cross protection phenomenon. Sequential deletion mutants of viral coat protein genes may help delineate this function. Experiments to date suggest that cross protection requires either a coat protein-mRNA association or a mRNA-replicase association. The mechanism may be better understood if more information regarding the assembly and disassembly of plant viruses is known. A coat protein-mRNA association may interfere either spatially or biochemically with the ribosome binding to mRNA for translation, or with the replicase binding for viral mRNA replication. Valle *et al.* (1988) noted that cross protection experiments to date integrate both the coat protein gene and the 3' non-coding region. They also suggest that it cannot be concluded whether cross protection is due to either coat protein expression, or to the 3'-termini of the genomes which presumably harbour the replicase binding site, or to both. This suggests the possibility of the replicase binding site being important in engineered as well as classical cross protection. Sequences relating to translation processes may also be important. Deletion mutants of the 3' and 5' untranslated regions of the coat protein gene may help define which processes are inhibiting infection by the challenging virus.

4.1.4. *Agrobacterium* MEDIATED PLANT TRANSFORMATION

The transfer and expression of foreign genes in plant cells has become a major tool for gene expression studies and for improving plant varieties of commercial interest (Herrera-Estrella and Simpson, 1988). The most successful plant transformation system is based on the Ti plasmid of *Agrobacterium tumefaciens*.

A.tumefaciens and *A. rhizogenes* are soil bacteria which induce the oncogenic phenotypes crown gall and hairy root disease respectively, at the wound site in dicotyledonous plants (Armitage *et al.*, 1988). The tumorous growth (or root proliferation) is initiated by the *Agrobacterium* but can continue growing without the bacteria. The tumorous tissue can grow axenically in tissue culture without a supply of exogenous auxin or cytokinin. These tissues also synthesise novel amino acids and sugar derivatives which are collectively termed 'opines'. Opines provide the bacteria with carbon and nitrogen sources (Armitage *et al.*, 1988) and the type of opine synthesised is dependent on the *Agrobacterium* strain. The genes responsible for tumour induction and opine synthesis are encoded by the tumour-inducing plasmid (Ti plasmid) of *A. tumefaciens* (or Ri plasmid in *A. rhizogenes*). The Ti plasmids from different strains of *Agrobacterium* have four regions of homology: the 'transfer' DNA (T-DNA) and 'vir' regions are associated with tumour formation and T-DNA transfer, and the two other regions function in the conjugative transfer and replication maintenance of the plasmid in *Agrobacterium* (Armitage *et al.*, 1988). The T-DNA is delineated by a pair of imperfect 25 base pair terminal repeat sequences (Wang *et al.*, 1984). An *Agrobacterium*-encoded endonuclease activity nicks these flanking repeats to produce a single stranded linear molecule of T-DNA which is integrated into the plant nuclear genome by a mechanism analogous to bacterial conjugation (Thomashow *et al.*, 1980). The T-DNA is believed to be transcribed to produce poly(A) mRNAs. Some of these overcome the normal regulation of the plant's metabolic pathways involved in the synthesis of phytohormones to produce the oncogenic phenotype.

The natural ability of *Agrobacterium* to transfer defined sections of DNA into plants has been exploited to construct a number of plant transformation vectors. This natural gene transfer system requires several modifications before it can be used as an efficient gene-vector system for producing transformed plants. First, the oncogenic genes are deleted from the T-DNA region and may be replaced with foreign DNA (Armitage *et al.*, 1988). Secondly, as the oncogenic growth can no longer be used to select transformed plant cells, a dominant marker gene is needed to confirm the integration and expression of the modified T-DNA. These selectable functions on most transformation vectors are prokaryote antibiotic resistant enzymes which are engineered to be expressed constitutively in plant cells. The coding region of these markers are normally fused with a promoter and polyadenylation signal. Most common is the kan^R (NPT II) gene (Herrera-Esterella *et al.*, 1983; Bevan *et al.*, 1985). NPT II phosphorylates and therefore inactivates kanamycin which is normally toxic to plant cells. The sensitivity of control plant tissue to kanamycin is 5-35 $\mu g\ ml^{-1}$, whereas resistant levels afforded by the construction are in the

order of 300-1000 μg ml^{-1} (Draper *et al.*, 1988). An alternative is an herbicide resistant marker (Shah *et al.*, 1986). Markers used to screen for transgenics include the Ti plasmid genes octopine synthase and nopaline synthase (Horsch *et al.*, 1984; Fraley *et al.*, 1983; Bevan *et al.*, 1983), and the enzyme chloramphenicol transacetylase. These are detectable by simple assays.

The basic components of a transformation vector are the 25 base pair T-DNA border sequences and a full complement of virulence genes, which may be present either on the same plasmid (cis) or on a different plasmid (trans) (Bevan, 1984; Hoekema *et al.*, 1983; Klee *et al.*, 1985). Trans, or binary vectors, are based on plasmids that replicate in both *E. coli* and *Agrobacterium*, and they contain the T-DNA border sequences. These are transferred by a tri-parental mating (van Haute *et al.*, 1983), or by transformation, into *Agrobacterium* containing a Ti plasmid which encodes the 'vir' genes but no T-DNA sequences. The helper functions of a third plasmid, for example pRK2013 (Ditta *et al.*, 1980), are required to facilitate the transfer.

A number of expression vectors have been designed which encode both markers for the direct selection of transformed plant cells, and border sequences which flank a multiple cloning site. This allows the easy creation of transcriptional fusions with promoters known to function in plant cells. The development of non-oncogenic, or disarmed vectors has led to the expression of genes in plant tissue cultures (Herrera-Esterella *et al.*, 1983; Bevan *et al.*, 1983; Fraley *et al*, 1983) and intact plants (Horsch *et al.*, 1985). Examples of the transcriptional regulatory sequences used to express these foreign genes include regions derived from nopaline synthase (Herrera-Esterella *et al.*, 1983; Bevan *et al.*, 1983) and octopine synthase genes, and from the nuclear encoded plant gene for the small sub-unit of ribulose bisphosphate carboxy oxygenase (Rubisco) (Herrera-Esterella *et al.*, 1984; Broglie *et al.*, 1984). Another source of gene regulatory elements is from the double stranded DNA genome of CaMV. Bevan *et al.* (1985) first used the highly active 35S promoter of CaMV to construct a hybrid gene consisting of the 35S promoter and cDNA from TMV encoding the coat protein gene.

Genetic transformation systems using *Agrobacterium* and components of TI and Ri plasmids have been developed for relatively few plant varieties. Most fundamental research on plant gene expression and control using chimaeric or modified genes, has been conducted in *Nicotiana tabacum*. This reflects the ease with which tobacco can be manipulated in culture and regenerated into whole plants (Draper *et al.*, 1988). Another model plant system used for transformation experiments is petunia. It is noteworthy that the alternatives are almost without exception *Solanaceous* species. Inoculation with the *Agrobacterium* species is commonly via the leaf disk protocol outlined by Horsch *et al.* (1985) .

4.1.5. MOLECULAR ANALYSIS OF TRANSFORMANTS

Following the transformation of plant tissue with a foreign gene and subsequent selection of transformants by antibiotic or herbicide resistance, the gene sequences must be screened for their presence and expression. It is possible to detect the presence of the gene in the

transformant and to determine the genomic organisation by Southern blotting (Southern, 1975). Qualitative studies of transcriptional activity and the presence of mRNA species can be estimated by Northern blotting (Alwine *et al.*, 1977). The expression and accumulation of protein can be monitored by Western blotting (Towbin *et al.*, 1979).

4.2. MATERIALS AND METHODS

4.2.1. CONSTRUCTION OF CHIMAERIC GENES

Transformation vectors containing transcriptional fusions of the PVY^N coat protein gene were constructed and subsequently inserted by triparental matings into an *A. tumefaciens* strain LBA4404 and *A. rhizogenes* strain A4T (G. Timmerman, CRD, DSIR, Private Bag, Christchurch).

The strategy for the vector constructions is outlined in Fig. 4.1. The cDNA containing the viral coat protein gene was excised from pVYN27 by digestion with *Cla*I and *Ssp*I restriction endonucleases. Complementary synthetic oligonucleotides were constructed (J. Cutfield, Biochemistry Department, Otago University, Dunedin) encoding an *Xba*I recognition site, AUG start codon, the nucleotide sequence for the three missing amino-terminal amino acids (alanine, asparagine, aspartic acid) plus the two amino acids removed by *Cla*I digestion (threonine and isoleucine) and a *Cla*I recognition site. The sequence postulated by Kozak (1986) as optimal for ribosome binding and translation initiation was also included. The attenuated coat protein sequence and the oligonucleotide were inserted into two expression cassettes on plasmids pCGN986 and pCGN46 to create chimaeric genes. pCGN986 (amp^R) contains the 35S promoter from cauliflower mosaic virus (CaMV), and pCGN46 (amp^R) contains the mannopine synthase promoter. Both vectors have polyadenylation signals derived from the Ti plasmid. Recombinant colonies were selected for ampicillin resistance and were transformed into *E. coli* strain C2110 containing the binary vector pCGN587. Vector pCGN587 is $chlor^R$, contains the left and right borders of the T-DNA region from *A. tumefaciens*, and encodes a chimaeric gene consisting of the octopine synthase (OCS) promoter and 5'-neomycin phosphotransferase (NPT II) gene. Strain C2110 is pol A defective. Co-integrates were selected for resistance to the appropriate bacterial antibiotics (amp^R, $chlor^R$). These were mobilised into *A. tumefaciens* strain LBA4404 (Hoekema *et al.*, 1983) containing pALA4404 (vir region) and *A. rhizogenes* strain A4T (Moore *et al.*, 1979) (hairy root strain) by a tri-parental mating using the mobilisation functions carried on pRK2013 (Ditta *et al.*, 1980).

After incubation on LB media to allow plasmid transfer, transconjugants were selected for *Agrobacterium* strains by growth on minimal media, and for *Agrobacterium* strains containing the co-integrate binary vectors by $carbenicillin^R$ and $chloramphenicol^R$.

The vectors pCGN986, pCGN46 and pCGN587, and the *Agrobacterium* strains LBA4404 and A4T were donated by A. J. Conner, CRD, DSIR, Private Bag, Christchurch.

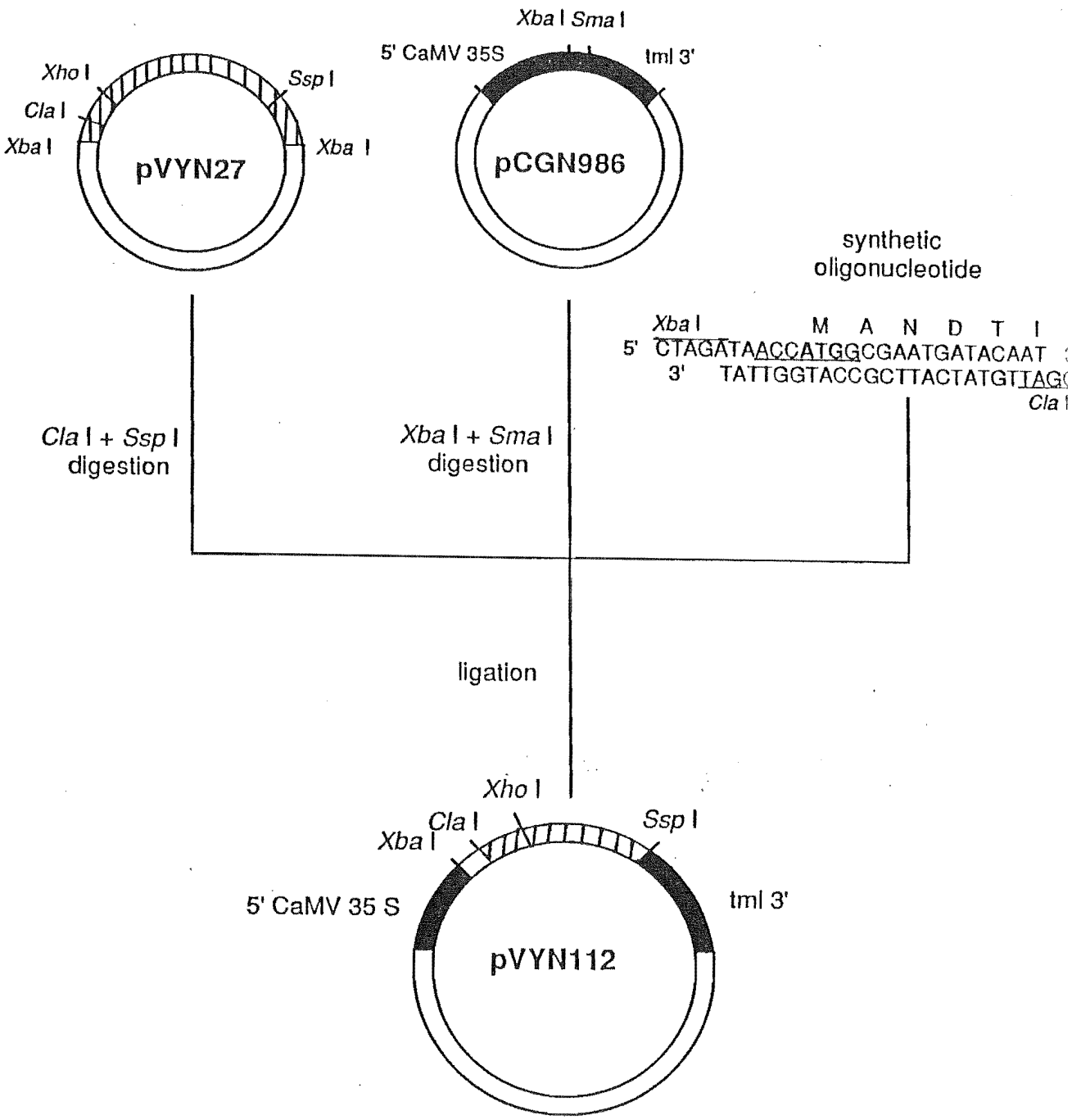


Figure 4.1. Strategy for the construction of chimaeric plasmids. The synthetic oligonucleotide duplex is shown as a DNA sequence. The sequence optimal for ribosome binding (Kozak, 1986) is underlined and includes an ATG start codon. The amino acid sequence coded for by the oligonucleotide is included. *Xba*I and *Cla*I recognition sites are encoded by the oligonucleotide.

4.2.2. PLANT TRANSFORMATION

Materials

All media were autoclaved at 121°C, 15 psi for 15 minutes after which filter sterilised antibiotics were added as required.

Table 4.1. *MS Medium (Murashige and Skoog, 1962).*

Component	Concentration (mg L^{-1})
CaCl2	440
NH4NO3	1650
KNO3	1900
KI	0.830
CoCl2.6H2O	0.025
KH2PO4	170
H3BO3	6.2
Na2MoO4.2H2O	0.25
MgSO4.4H2O	370
CuSO4.5H2O	22.3
ZnSO4.4H2O	0.025
FeSO4	8.6
Na2EDTA	27.85
	37.25

Adjust the pH to 5.8 before autoclaving.

RMNOP medium for callusing and regeneration (Maliga, 1984)

MS containing 1 mg L^{-1} thiamine-HCl, 100 mg L^{-1} myo-inositol, 1 mg L^{-1} 6-benzylaminopurine (BAP), 0.1 mg L^{-1} naphthaleneacetic acid (NAA), 3% sucrose, 0.7% agar, pH 5.8.

Potato Callus Medium (PCM) (Williams et al., 1988)

MS containing 1 mg L^{-1} thiamine-HCl, 100 mg L^{-1} myo-inositol, 40 mg L^{-1} ascorbic acid, 500 mg L^{-1} casein hydrolysate, 0.2 mg L^{-1} NAA, 2 mg L^{-1} BAP, 3% sucrose, 0.7% agar, pH 5.8.

Hairy Root Regeneration (HR) Medium (M. Witty, pers. comm.)

MS containing 5 mg L^{-1} gibberellic acid (GA_3), 1 mg L^{-1} indole acetic acid (IAA), 2 mg L^{-1} BAP, 3% sucrose, 0.7% agar, pH 5.8.

Table 4.2. *MG/L Broth (Garfinkel and Nester, 1980)*

Component	Concentration
mannitol	5.0 g L^{-1}
Na glutamate	1.15 g L^{-1}
bactotryptone (Difco)	5.0 g L^{-1}
yeast extract (Difco)	2.5 ml
10% MgSO4	2.5 ml
10% NaCl	1.0 ml
10% KH2PO4	2.5 ml
10-4% biotin	1.0 ml

Adjust the pH to 7.0 before autoclaving.

Methods

Axenic *Solanum tuberosum* cv. Iwa and *Nicotiana plumbaginifolia* plants, donated by A. J. Connor (CRD, DSIR, Private Bag, Christchurch), were maintained *in vitro* on half-strength MS salts (1/2 MS), containing 3% (w/v) sucrose and 0.7% (w/v) agar. Twenty mls of MG/L broth containing 70 mg L^{-1} chloramphenicol and 150 mg L^{-1} carbenicillin were inoculated with a single colony of *Agrobacterium* strains LBA4404 or A4T harbouring the chimaeric coat protein gene, and grown with agitation overnight at 28°C.

Tobacco

Six mm leaf disks from three month old, regenerated axenic plants were dipped for approximately 30 seconds in the overnight *Agrobacterium* suspension and blotted onto sterile filter paper. They were placed on RMNOP medium in 9 cm disposable Petri dishes and sealed with 'Gladwrap'. After two days the leaf disks were transferred to RMNOP containing the bacteriostatic antibiotic cefotaxime (300 mg L^{-1}) and a transformant selection agent, kanamycin (300 mg L^{-1}), for the initiation of callus cultures. Shoots were regenerated from callus on RMNOP containing antibiotics as above.

Potato

Six mm leaf disks were treated with *Agrobacterium* as for tobacco. The disks were cultured on a potato callusing medium (PCM) in disposable Petri dishes and sealed with 'Gladwrap'. After two days the disks were transferred to PCM containing 250 mg L^{-1} cefotaxime to inhibit overgrowth

of *Agrobacterium*. Four to five days later the disks were transferred to PCM containing 250 mg L^{-1} cefotaxime and 250 mg L^{-1} kanamycin to select for transformed colonies.

For the transformation of stem sections of potato by the *Agrobacterium* strain A4T, one cm long internodal stem sections from axenic potato plants were inoculated with the *Agrobacterium* suspension as described. The sections were then incubated on PCM as described for potato leaf disks. After three days the stems were transferred to PCM containing 300 mg L^{-1} cefotaxime and after four days were transferred to PCM containing 300 mg L^{-1} cefotaxime and 25 mg L^{-1} kanamycin. Shoots were regenerated from callus on HR regeneration medium containing 100 mg L^{-1} cefotaxime and 25 mg L^{-1} kanamycin.

All axenic plants and cultures were incubated under cold white fluorescent light ('Gro lux', Sylvania) at 25°C for a 16 hour day and 18°C for an 8 hour night. All tissue culture manipulations were performed aseptically in a Laminar flow cabinet (Gelman Sciences).

4.2.3. SELECTION OF TRANSFORMANTS

Transformants were selected for their resistance to kanamycin (kan^R). Calli and hairy roots which formed following transformation of the leaf disks and stems were transferred every one to two weeks to fresh selective media containing kanamycin and cefotaxime. Discrete colonies which produced large clumps of transformed callus were dissected out and the cut edges of the colonies were pressed back onto the medium to retain selective pressure for kanamycin resistance.

4.2.4. PLANT REGENERATION

Tobacco

Kanamycin resistant shoots developed on the RMNOP medium and were transferred to fresh RMNOP containing 300 mg L^{-1} kanamycin and 200 mg L^{-1} cefotaxime in 60 mm x 85 mm tissue culture vessels (Life Tech Lab.) to permit shoot expansion. Transformed shoots (> 0.5 cm) were excised from the callus clumps and trimmed leaving the apical bud and 2-3 leaves. These individual shoots were transferred to 1/2 MS containing 0.7% agar and 100 mg L^{-1} kanamycin in tissue culture vessels to initiate root development. After a root system began to develop individual plantlets were transferred to 1/2 MS medium without antibiotics. These were then potted into sterile potting mix (peat/sand mix, 60/40) and covered with an upturned tissue culture vessel to avoid dessication of the plants and to facilitate acclimatisation. One week later the covering containers were removed. The plantlets were maintained in a growth room at 22°C with a 16 hour day and an 8 hour night. When plants had 4-6 leaves, assays to detect the presence of virus specific RNA or viral encoded protein were performed.

4.2.5. PROTEIN ANALYSIS OF TRANSFORMED PLANTS

Protein Preparation

0.1 g of leaf tissue was homogenised in 800 μl TBS (20mM Tris-HCl pH 7.5, 500mM NaCl) in a tissue grinder. The homogenate was transferred to an eppendorf tube and cell debris pelleted by microfugation at 12,000 rpm for 10 minutes. Samples of 35 μl were diluted in 200 μl of TBS (approximately 450 μg protein) for slot blot analysis.

For Western blot analysis, 0.1 g of leaf tissue was homogenised in 800 μl of phosphate buffered saline, pH 7.3, (PBS) (4mM KH_2PO_4, 16mM Na_2HPO_4, 115mM NaCl) and 8 μl samples (approximately 100 μg protein) were used for analysis.

Protein Slot Blot Analysis (Biodot Manual, Biorad)

Four sheets of Whatman 3mm filter paper and a nitrocellulose membrane (NCM) ('Biotrace', Gelman) were pre-wet in TBS and placed on the slot blot apparatus (Biorad). The apparatus was assembled and operated as per the manufacturers instructions. 200 μl of TBS were pipetted into each slot and suction applied. 200 μl of diluted protein sample were then applied and sucked onto the membrane, followed by a wash with 200 μl of TBS.

The membrane was removed and 'blocked' with PBS containing 0.05% Tween 20 (Batteiger *et al.*, 1982) and 3% non-fat milk powder (Anchor) for two hours at 25^{o}C with agitation. The membrane was then drained and washed twice with PBS for 5 minutes and twice with PBS, 0.05% Tween 20 for 15 minutes. The NCM was sealed in a plastic bag containing the primary anti-PVY antibody (Boehringer mannheim) at a 1:4000 dilution in PBS, 0.05% Tween 20, 3% acetylated BSA. 0.1 mls of solution was added per cm^2 of membrane and was shaken vigorously overnight at 25^{o}C (Ey and Ashman, 1986).

Following three 5 minute washes in PBS, 0.05% Tween 20, the membrane was reacted with donkey anti-sheep antibody conjugated to alkaline phosphatase (Silenus) (secondary antibody) diluted to 1:8000 in PBS, 0.05% Tween 20, 3% BSA. The membrane was sealed in a plastic bag and shaken at 25^{o}C for two hours. It was then washed as before and incubated with a modification of the substrates originally described by M^{c}Gadey (1970). The substrate solution contained 20 μl 2M $MgCl_2$, 1.0 ml nitro blue tetrazolium (NBT) (0.1% solution in 0.15M veronal acetate, pH 9.6), 0.1 ml 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (5 mg ml^{-1} solution in dimethylformamide) and 9.0 mls of 0.15M veronyl acetate buffer, pH 9.6. The reaction mix was gently agitated and colour development proceeded in the dark at room temperature, until sufficient colour was observed (normally 10-15 minutes). The membrane was washed in dH_2O and allowed to air dry.

Western Blot Analysis (Towbin *et al.*, 1979)

Eight μl samples of crudely prepared leaf sap were added to 2 μl of 5x sample buffer (0.3M Tris-HCl pH 6.8, 10% SDS, 62.5% (w/v) glycerol, 3.5M β-mercaptoethanol, 0.004% BPB) and

denatured at 100°C for 90 seconds. A mini SDS-polyacrylamide gel containing 12.5% bis : acrylamide (0.8 : 29.2), 0.1% SDS, 0.1% APS, 4 μl TEMED in 0.37M Tris-HCl pH 8.8 was prepared (Laemmli, 1970). The stacking gel contained 3% acrylamide, 0.1% SDS, 0.04% APS, 7.5 μl TEMED in 0.125M Tris-HCL pH 6.8 and the electrophoresis buffer contained 30mM Tris base pH 8.3, 0.240M glycine and 0.125% SDS. Samples were run at 60 mA through the stacking gel and 30 mA through the separating gel. Following electrophoresis, the gel was removed and soaked in ice cold transfer buffer (20mM Tris-HCl, 192mM glycine, 20% (v/v) methanol) for 15 minutes. The protein bands were transferred to NCM (Gelman) in ice cold transfer buffer for one hour at 250 mA and 100 V following the manufacturer's instructions (Biorad).

After transfer, the membrane was removed and 'blocked' for two hours at 25°C with PBS containing 0.05% Tween 20, 3% non-fat milk powder. The method for the reaction of the primary and secondary antibodies, and colourimetric development continued as described for protein slot blot analysis.

The efficiency of transfer of the protein bands was verified by staining the gel with Coomassie R-250 (section 2.2.8.). Low molecular weight protein markers (Biorad) were prepared according to the manufacturer's instructions. Ten μl were loaded onto the gel, electrophoresed and transferred onto NCM with the sample lanes. The strip of NCM containing the marker lane was removed after blocking and stained with 10.0 ml of Coomassie G-250 (Bradford, 1976) for 30 minutes. The markers were destained for 2-3 hours in 5% glacial acetic acid and finally rinsed in dH_2O and air dried.

4.2.6. PROBE PREPARATION

^{32}P-labelled DNA probe, synthesized using the Multi-Prime DNA Labelling system (Amersham), was used to probe Northern and Southern blots, and DNA and RNA slot blots.

The 956 base pair *Cla*I restriction endonuclease fragment of pVYN27 was isolated and purified as in section 3.2.1. (Plate 4.1.). Twenty five ng of *Cla*I fragment were denatured at 100°C for five minutes and cooled on ice. Labelled fragments were generated in a reaction mix containing 30 μCi of ^{32}P-dCTP (3000 Ci mmol^{-1}), according to the manufacturer's instructions. The reaction was incubated for 5 hours at room temperature. One μl of reaction was diluted in 400 μl of 10% TCA, 10mM Na pyrophosphate and 50 μl denatured herring sperm DNA (1 mg ml^{-1}). The specific activity of the labelled fragment was calculated from TCA precipitable counts (section 2.2.11.). The probe was stored at -20°C after the addition of EDTA to 20mM.

A 1.25 kilobase pair NPT II DNA fragment was supplied by G. Timmerman (CRD, DSIR, Private Bag, Christchurch) and was labelled using the same reaction.

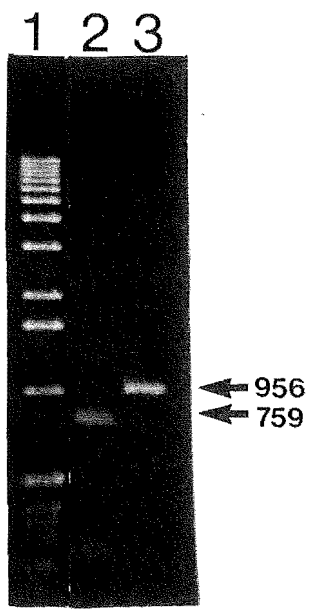


Plate 4.1. Electrophoresis of the *Cla*I fragment isolated from pVYN27. **Lane 1.** 1 μg BRL 1kb DNA ladder. **2.** 1 μg *Cla*I fragment digested with 1 unit *Xho*I for an hour at 37^{O}C. *Xho*I cleaves the *Cla*I fragment once, generating two fragments of 759 and 197 (not visible) base pairs, respectively. **3.** 1 μg *Cla*I fragment isolated from pVYN27.

4.2.7. DNA ANALYSIS OF TRANSFORMED PLANTS

DNA Preparation

DNA was isolated from transgenic tobacco plants by the method of Dellaporta *et al.* (1983). DNA was isolated from potato hairy roots by the same method, with 1% ascorbic acid added to the initial extraction buffer. 0.8 g of leaf tissue was frozen in liquid nitrogen and ground to a fine powder with a mortar and pestle. The powder was transferred to a 50.0 ml polypropylene centrifuge tube and 15.0 ml of extraction buffer were added (100mM Tris-HCl pH 8.0, 50mM EDTA, 500mM NaCl, 10mM β-mercaptoethanol). One ml of 20% SDS was then added, mixed by vigorous shaking and the homogenate was incubated on ice for 20 minutes. The tube was centrifuged at 15,000 rpm in an IEC centrifuge rotor for 25 minutes and the supernatant collected and filtered through miracloth. Ten ml of isopropanol were added to the supernatant, mixed and incubated at -20^{o}C for 30 minutes. The DNA was collected by centrifugation at 15,000 rpm for 15 minutes. The pellet was redissolved in 700 μl of 50mM Tris-HCl pH 8.0, 10mM EDTA and transferred to an eppendorf tube. Insoluble debris was removed by spinning at 13,000 rpm for 10 minutes. 75 μl of 3M Na acetate and 500 μl of isopropanol were added to the supernatants, mixed well and the DNA pelleted for 30 seconds in a microfuge. The pellet was washed with 70% ethanol, allowed to dry and resuspended in 100 μl of TE (10mM Tris-HCl pH 8.0, 1mM EDTA).

The DNA concentrations of the resuspended pellets were estimated by spectrophotometric readings at 260 nm (Maniatis *et al.*, 1982).

DNA Slot Blot Analysis

DNA samples were analysed by a modification to the dot blot procedure outlined in the Genescreen Plus manual (DuPont). Four sheets of Whatman 3mm filter paper and a piece of nylon membrane (Genescreen Plus) were soaked for 30 minutes in 0.4M Tris-HCl pH 7.5 and fitted onto the blot manifold (Biorad). Five to ten μg of sample DNA were denatured in 0.25N NaOH for 10 minutes and chilled on ice. The DNA was diluted to 200 μl in 0.125N NaOH, 0.125x SSC (0.02M NaCl, 0.002M Na citrate) and applied in duplicate to the wells of the manifold. The solutions were sucked through the wells onto the nylon after 30 minutes. The membrane was then removed and air dried.

The membrane was pre-hybridised in 6x SSC (0.9M NaCl, 0.09M Na citrate), 1% SDS, 10% dextran sulphate and 300 μg ml^{-1} denatured herring sperm DNA in a plastic bag. Pre-hybridisation solution was added to 0.1 ml cm^{-2} and incubated between 30 minutes and overnight at 65^{o}C with constant agitation. ^{32}P-labelled *Cla*I fragment was denatured at 100^{o}C for 5 minutes and added to 2.0 ml of freshly made pre-hybridisation solution. This was then added to the bag containing the pre-hybridisation buffer and the nylon, and resealed. The bag was agitated at 65^{o}C for 6-24 hours. Following hybridisation, the membrane was removed and washed twice in 100 ml 2x SSC at room temperature for 5 minutes, twice in 200 ml 2x SSC, 1.0%

SDS at 65^{o}C for 30 minutes and twice in 100 ml 0.1x SSC at room temperature for 30 minutes. The membrane was sealed in a bag and exposed to X-ray film ('Cronex 4', DuPont) at -80^{o}C, using two intensifying screens.

Southern Blot Analysis (Southern, 1975)

Fifteen μg of transformed plant DNA were digested overnight at 37^{o}C with 30 units of *BamHI* restriction endonuclease in a 30 μl reaction containing 1.0 μg of heat treated RNase and 0.5μM spermidine. Ten μl of 5x sample buffer (50% (w/v) glycerol, 0.25% BPB in 5x TBE) were added to each sample and loaded onto a 0.7% agarose horizontal gel in 1x TBE (89mM Tris-borate, 89mM boric acid, 8mM EDTA). One μg of 1 kb DNA markers (BRL) was also loaded and samples were electrophoresed overnight at 30 V in 1x TBE buffer. Following electrophoresis, the gel was lightly stained in a 0.5 μg ml^{-1} ethidium bromide solution. The DNA was visualised on a short wave ultraviolet transilluminator and a photograph was taken.

The blotting and hybridisation protocols used were modifications of those described in the Genescreen Plus manual (Biorad). The gel was gently agitated in 0.4N NaOH, 0.6M NaCl for 30 minutes at room temperature to denature the DNA. It was then neutralised in 1.5M NaCl, 0.5M Tris-HCl pH 7.5 for 30 minutes at room temperature. Nylon membrane (Genescreen Plus) was cut to the size of the gel, wet in dH_2O and soaked in 10x SSC (1.5M NaCl, 0.15M Na citrate) for 15 minutes at room temperature. The blot was assembled and the DNA transferred by capillary action in 10x SSC, overnight. After disassembling the blot, the membrane was immersed in an excess of 0.4N NaOH for 60 seconds to ensure complete denaturation of the immobilised DNA. The drained membrane was immersed in an excess of 0.2M Tris-HCl pH 7.5, 2x SSC, drained and allowed to air dry. Pre-hybridisation, hybridisation, washing and autoradiography were the same as described for DNA slot blot analysis.

4.2.8. RNA ANALYSIS OF TRANSFORMED PLANTS

RNA Preparation

RNA was isolated from transgenic tobacco plants by the method of Habili *et al.* (1987). 0.1 g of leaf tissue was homogenised in 1.0 ml of AMES buffer (0.5M NaOAc pH 6.0, 10mM $MgCl_2$, 3% (w/v) SDS, 20% ethanol) and 1.0 ml of water saturated phenol. The mixture was spun for 10 minutes at 13,000 rpm in a microfuge and the supernatant collected. The supernatant was extracted once with an equal volume of chloroform (chloroform:iso-amyl alcohol, 24:1) and the phases separated by microfuging at 13,000 rpm for 10 minutes. The aqueous phase was collected and the DNA was ethanol precipitated. The dried RNA pellet was resuspended in 100 μl TE.

The RNA concentrations of the resuspended pellets were estimated by spectrophotometric

readings at 260 nm (Maniatis *et al.*, 1982). All manipulations for preparing RNA were done under RNase free conditions (section 2.2.15).

Northern Blot Analysis (Alwine *et al.*, 1977)

The preparation of gels and samples for Northern blotting were by the procedure of Fourney *et al.* (1988). Approximately 50 μg of RNA isolated from transgenic tobacco were resuspended in 5 μl of 25mM EDTA, 0.1% SDS. Twenty five μl of electrophoresis buffer (0.75 ml deionised formamide, 150 μl 10x MOPS (200mM MOPS [3-(N-morpholino) propanesulfonic acid] pH 7.0, 500mM Na acetate, 100mM EDTA), 240 μl formaldehyde, 100 μl dH2O, 100 μl glycerol, 80 μl of 10% (w/v) BPB) were added to each sample, which was denatured at 65^{o}C for 15 minutes. One μl of ethidium bromide (1.0 mg ml^{-1}) was introduced into each sample before being loaded onto the gel. One μg of 1kb DNA markers (BRL) was also electrophoresed.

A 1% agarose horizontal gel containing 1x MOPS and 5.1 ml of 37% formaldehyde was made to 30.0 ml, poured and allowed to set. Samples were electrophoresed overnight at 30 V at room temperature in 1x MOPS buffer. The RNA was visualised on a short wave transilluminator and photographed. The gel was prepared for transfer by agitation in 10x SSC at room temperature for two 20 minute periods. The nylon membrane (Genescreen Plus) was equilibrated by wetting it in dH_2O followed by soaking in 10x SSC for 15 minutes. RNA was transferred to the membrane overnight in 10x SSC by capillary action.

All subsequent manipulations of the nylon were modified methods from the Genescreen Plus manual (DuPont). The membrane was removed from the gel, rinsed in 2x SSC to remove residual agarose and air dried. It was then baked for two hours at 80^{o}C to reverse the formaldehyde reaction. The membrane was pre-hybridised in 0.1 ml cm^{-2} of 1% SDS, 50% deionised formamide, 6x SSC, 10% dextran sulphate and 300 μg ml^{-1} denatured herring sperm DNA. The solution was added to the membrane in a plastic bag, sealed and agitated between 30 minutes and overnight at 42^{o}C. ^{32}P-labelled probe was diluted in 2.0 ml of freshly prepared pre-hybridisation solution and added to the membrane in the sealed bag. Hybridisation continued overnight at 42^{o}C with constant agitation. The membrane was removed and washed with agitation, twice in 100 ml 2x SSC at room temperature for 5 minutes, twice in 200 ml 2x SSC, 1% SDS at 60^{o}C for 30 minutes and twice in 100 ml 0.1x SSC at room temperature for 30 minutes. The nylon membrane was sealed in a plastic bag and the autoradiograph exposed and developed (Maniatis *et al.*, 1982).

4.2.9. SEED GERMINATION

Agar plates containing 1/2 MS and 0.6% agar were poured. Half of these contained 400 mg L^{-1} kanamycin. Self fertilised seed was collected from transgenic tobacco plants and soaked overnight in gibberellic acid (1.4 x10^{-3}M GA_3). They were surface sterilised in a bleach solution (1% NaClO) for 10 minutes and rinsed in dH_2O. Fifty seeds from a single transgenic plant were

transferred in duplicate to Petri dishes containing media with and without kanamycin. Seeds from non-transformed *in vitro* plants and commercial seed were plated as controls. The number of seeds germinating was tabulated. Chi square tests for goodness of fit and tests of association, as appropriate, were used to test differences between germination percentages, and for a 3:1 segregation ratio.

4.2.10. CROSS PROTECTION EXPERIMENT

A total of 56 plants bearing 3-4 leaves were inoculated on a single leaf with a 0.5 μg ml^{-1} suspension of PVY^N. Replicate *N. plumbaginifolia* plants with integrated chimaeric genes from plasmids pVYN112 (in *Agrobacterium* strains LBA4404 and A4T), pVYN53 (in strain A4T) and pVYN51 (in strain A4T) were inoculated. Untransformed plants and uninfected plants served as controls. The plants were placed in a controlled growth room and monitored daily.

4.3. EXPERIMENTS AND RESULTS

4.3.1. VECTOR CONSTRUCTS

Vectors containing transcriptional fusions of the PVY^N gene and either the mannopine synthase or CaMV 35S promoter were constructed as shown in Table 4.3. These plasmids were recombined into the binary plasmid pCGN587. Co-integrates which were $chlor^R$ and amp^R were selected. These were purified and mapped for diagnostic fragments to confirm the construct and orientation of the chimaeric gene. (For the remainder of this report the binary vectors formed after the recombination of plasmids carrying the chimaeric PVY^N coat protein gene constructs and pCGN587, will be referred to by the name of the initial chimaeric plasmid, for example pVYN112.) Following mobilisation of the co-integrates into two strains of *Agrobacterium* (A4T and LBA4404), transconjugants with the appropriate phenotype ($carb^R$, $chlor^R$) were selected and streaked out onto fresh media.

4.3.2. PLANT TRANSFORMATION

Tobacco

Transformation of *N. plumbaginifolia* leaf disks was successfully achieved using the following vectors (Table 4.4.) in the strains indicated (Plate 4.2.).

Table 4.4. Transformation of *N. plumbaginifolia* leaf disks with chimaeric plasmids.

Strain[a]	Binary vector[b]
A4T	pVYN42
A4T	pVYN41
A4T	pVYN53
A4T	pVYN51
A4T	pVYN112
A4T	pVYN113
A4T	pVYN46
A4T	-
LBA4404	pVYN41
LBA4404	pVYN42
LBA4404	pVYN112
LBA4404	pVYN113
LBA4404	-

a. *Agrobacterium* strain into which the binary vectors were successfully mobilised.

b. Binary vectors containing the chimaeric constructs defined in Table 4.3.

Table 4.3. Plasmids carrying chimaeric PVY^N coat protein gene constructs

Plasmid	Promoter	PVY^N C.P. Construct[b]	Phenotype[c]
pCGN46	mannopine synthase	No insert	Amp^R
pCGN986	35S CaMV	No insert	Amp^R
pCGN587[d]	-	No insert	$Chlor^R$
pVYN27	-	PVY^N C.P. gene	Amp^R
pVYN41[a]	35S CaMV	PVY^N C.P. gene (anti)	Amp^R, $Chlor^R$
pVYN42[a]	35S CaMV	PVY^N C.P. gene (syn)	Amp^R, $Chlor^R$
pVYN51[a]	mannopine synthase	PVY^N C.P. gene (anti)	Amp^R, $Chlor^R$
pVYN53[a]	mannopine synthase	PVY^N C.P. gene (syn)	Amp^R, $Chlor^R$
pVYN112[a]	35S CaMV	PVY^N C.P. gene (syn) +	Amp^R, $Chlor^R$
pVYN113[a]	35S CaMV	synthetic oligonucleotide	Amp^R, $Chlor^R$

a. Chimaeric vectors containing transcription fusions of the PVY^N coat protein gene.

b. The PVY^N coat protein (C.P.) gene was cloned into the chimaeric vector in either a sense (syn) or anti-sense (anti-) direction. The synthetic oligonucleotide was cloned into pVYN112 and pVYN113 only. Vectors pCGN46, PCGN587 and pCGN986 contained no inserts and were controls for the presence of the PVY^N C.P. gene.

c. Co-integrates formed after recombination of the chimaeric plasmid and pCGN587 were Amp^R, $Chlor^R$. These binary plasmids are subsequently referred to by the names listed in column 1. The *Agrobacterium* transconjugants were $Carb^R$, $Chlor^R$. Transformed plants were kan^R.

d. Binary vector with broad host range.

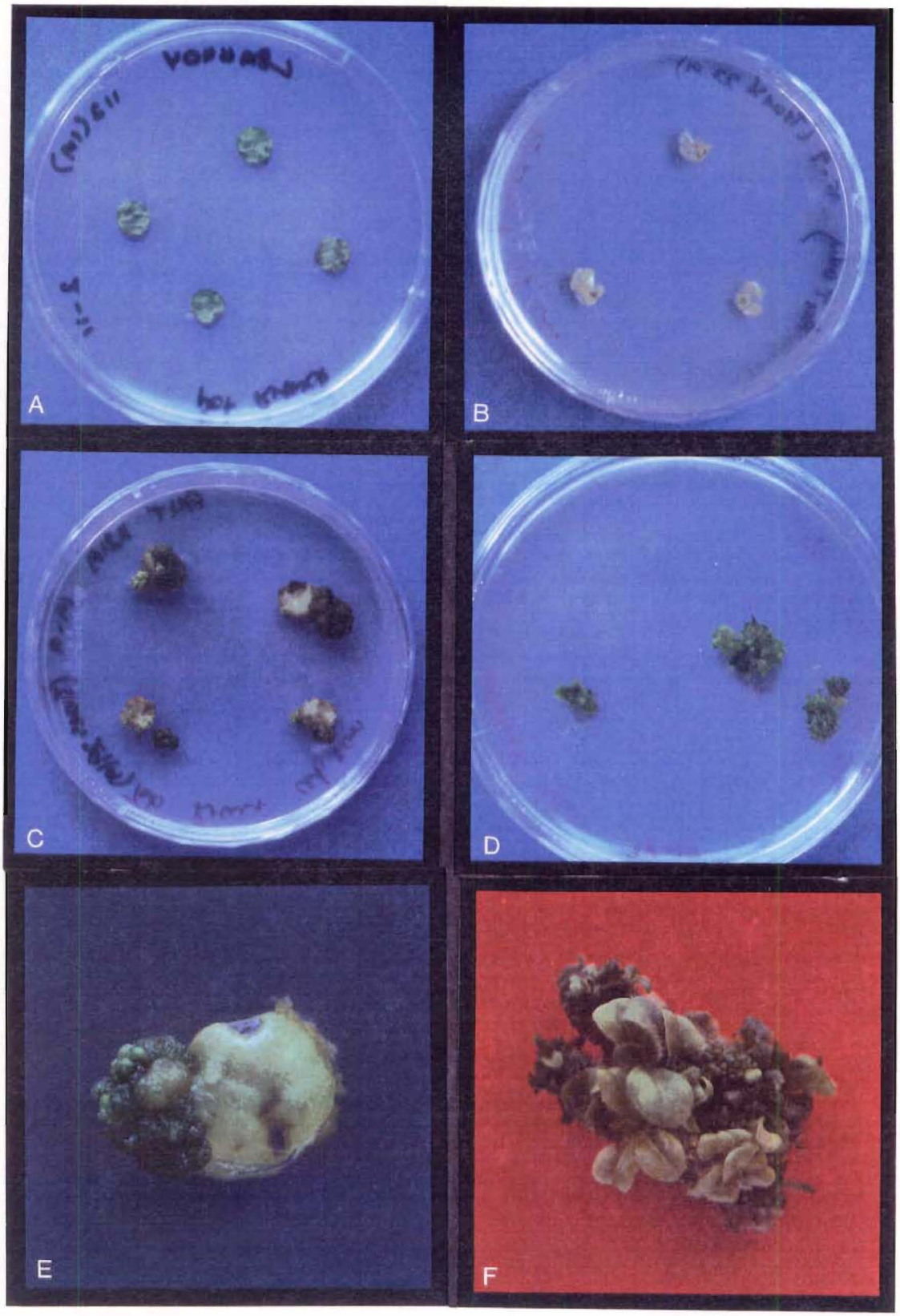


Plate 4.2. Transformation of *N. plumbaginifolia* leaf disks with *Agrobacterium* strains harbouring chimaeric plasmids, and placed on cef^+, kan^+ media. **(A)** Day 1. **(C)** Two to three weeks later. **(D)** Transformed calli formed from discrete colonies were dissected from the leaf disk. **(E)** Transformed callus forming on a leaf disk. **(F)** Kanamycin resistant shoots developing from transformed calli after 6-7 weeks. **(B)** Leaf disks inoculated with *Agrobacterium* harbouring no chimaeric plasmid and placed on cef^+, kan^+ media. After 1 week.

Control leaf disks were transformed with *Agrobacterium* A4T and LBA4404 strains which harboured none of the chimaeric binary vectors. These bleached within one to two weeks when plated onto media containing cefotaxime and kanamycin (cef+, kan+) (Plate 4.2.(B)) which confirmed that kan^R was conferred to developing callus by the co-integrates. The vector pCGN986 was never successfully mobilised into either *Agrobacterium* strain and therefore could not be used as a control for subsequent expression experiments. Transformation with *Agrobacterium* strain A4T containing pCGN46 was successful. Transformation of tobacco leaf disks with either strain containing pCGN587 was not attained.

Potato

Potato leaf disks were successfully transformed with a number of the binary vectors in *Agrobacterium* strains LBA4404 and A4T. Callus formed but no shoot regeneration was observed for any of the transformed colonies.

As an alternative to leaf disk transformation, production of callus and regeneration of whole plants was attempted from hairy roots infected with the *Agrobacterium rhizogenes* strain containing chimaeric binary vectors. Stem sections were successfully transformed with the vectors pVYN41, pVYN42, pVYN51, pVYN53, pVYN112 and pVYN113 and the control vectors pCGN986 and pCGN587 in the *Agrobacterium* strain A4T. Prolific growth of hairy roots on kan^+ PCM medium was observed from all stems infected with non-control vectors (Plates 4.3.). Upon transfer to HR regeneration medium, cell proliferation occurred and the roots began to produce callus. Unfortunately, the time required for sufficient callus to form and for shoots to be produced amounted to months. Because of this, DNA was prepared from the roots themselves and this was tested for the presence of the coat protein gene. Callusing roots were transferred to CRD, DSIR, Private Bag, Christchurch for maintenance and regeneration of whole plants.

4.3.3. PLANT REGENERATION

Kanamycin resistant callus was observed to form on *Nicotiana plumbaginifolia* leaf disks after 2-3 weeks culture and was well developed after 5-6 weeks. RMNOP medium induces organogenesis as well as callusing, and shoots began to appear after a further 3-4 weeks. Morphologically normal shoots (approximately one cm long) were dissected from the callus clumps and transferred to 1/2 MS for rooting. No organic compounds were added to the medium, thereby promoting photoautotrophic growth. After a root system was established (2-3 weeks), individual plantlets were transferred to fresh 1/2 MS for a further 5-6 weeks. Plantlets were then transferred out of culture and into a peat/sand potting mix. Seventy five kanamycin resistant tobacco plants, transformed with a variety of chimaeric genes were potted out and tested for coat protein

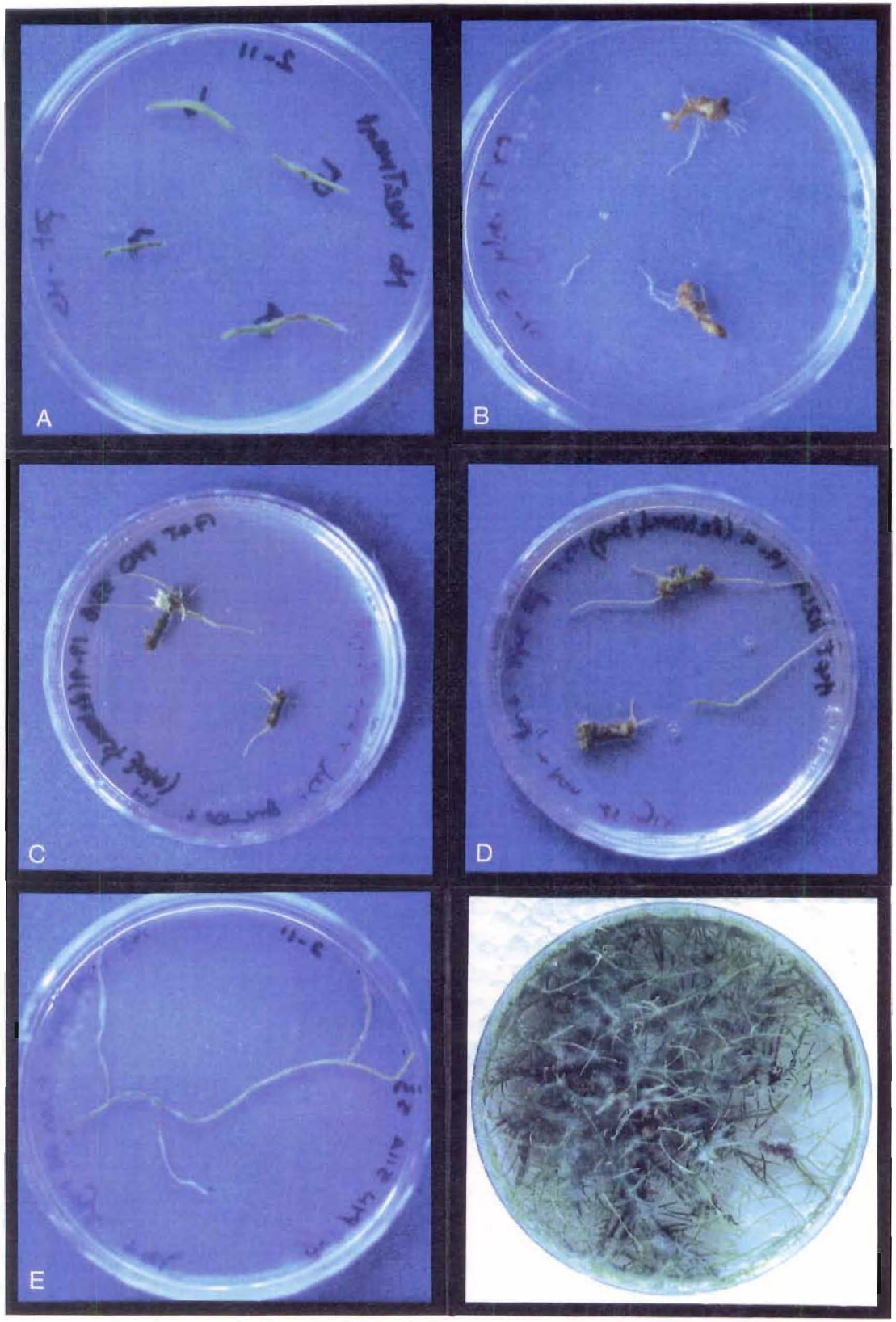


Plate 4.3. Transformation of *S. tuberosum* stem explants with *Agrobacterium* strain A4T harbouring chimaeric plasmids, and placed on cef$^+$, kan$^+$ media. **(A)** Day 1. **(C)** Two weeks later. **(D)** Three weeks later. **(E)** Single roots were dissected from the explants. **(F)** Kanamycin resistant hairy roots after 6-7 weeks. **(B)** Stem explants inoculated with *Agrobacterium* harbouring no chimaeric plasmid, after 1 week on media containing cef+, kan+.

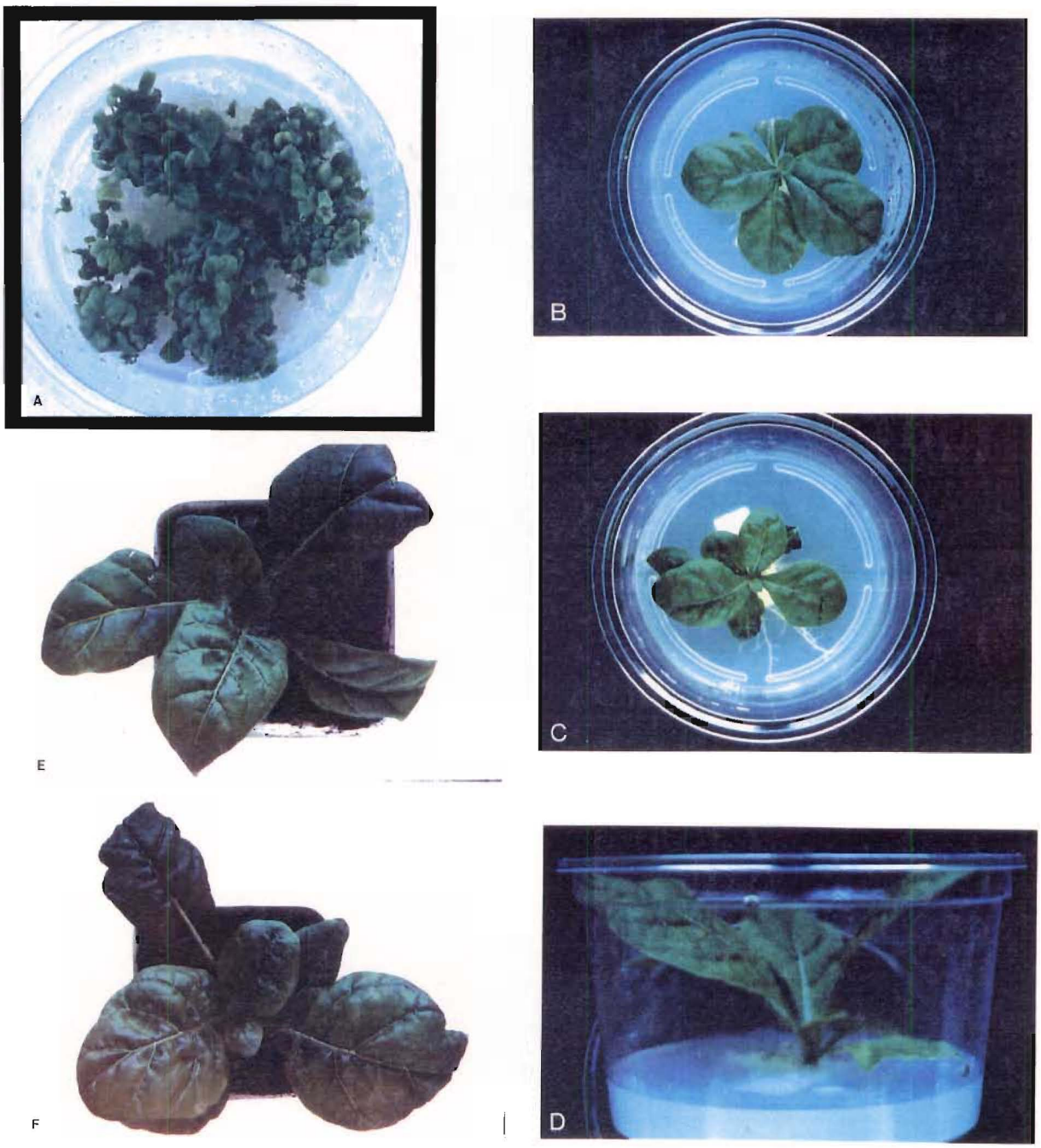


Plate 4.4. Regeneration of transformed *N. plumbaginifolia*. **(A)** Calli after transfer to tissue culture vessels. **(B)** Single transformed shoots after transfer to root initiating media. **(C)** Root development after 2-3 weeks. **(D)** Plantlet. **(E)** Transformed plantlet in potting mix. **(F)** Untransformed plantlet in potting mix.

expression and for the presence of the coat protein gene. Plate 4.4. illustrates some of the different developmental stages of tobacco plant regeneration. Regenerated kan^R plant cells developed into morphologically normal tobacco plants. No virus-like symptoms or other abnormalities in growth, flowering or seed production were observed.

Whole plants were not regenerated from potato leaf disks or from potato hairy roots.

4.3.4. PROTEIN ANALYSIS OF TRANSFORMED PLANTS

A preliminary Western blot and protein slot blot experiment was undertaken to test the primary anti-PVY antiserum. Protein samples were prepared from uninfected, PVY^N- and PVY^C-infected tobacco leaf tissue as described. For the Western blot experiment approximately 63 μg of total plant protein were loaded onto a 12.5% Laemmli gel. Five μg of purified pea seed-borne mosaic virus (PSbMV) was also loaded to test the specificity of the antiserum. For the protein slot blot experiment approximately 2.5 mg of total plant protein from uninfected, PVY^N and PVY^C infected tobacco sap were applied to the membrane. A five μg sample of PSbMV was diluted in 200 μl and was also applied to the membrane using the Biorad manifold. In both analyses the primary anti-PVY antiserum was conjugated to alkaline phosphatase and was used at a 1:2000 dilution to probe the membranes. For the Western blot, transfer of protein from the gel to the membrane was shown to be complete by staining the gel with Coomassie blue R-250. Plate 4.5. shows the resulting Western blot. Lanes 2 and 3 loaded with PVY^C and PVY^N infected sap each revealed a band which migrated 2.4 cm, corresponding to a molecular weight of 33 by comparison with the standards. No bands were observed in lane 1 containing the uninfected sap sample. Lane 4 was loaded with PSbMV and showed a positive signal, but some non-specific binding was evident in the lane. This was possibly due to the antiserum being too concentrated or to the sample being overloaded. The protein slot blot (Plate 4.6.) indicated positive signals for PVY^N and PVY^C infected tobacco sap and for PSbMV. All subsequent protein blots used a primary antibody and a secondary antibody at the adjusted dilutions of 1:4000 and 1:8000, respectively.

To test the sensitivity of the slot blot procedure, a dilution series of purified PVY^N (1 mg ml^{-1}) was applied to the nylon membrane and assayed with primary and secondary antibodies at the stated dilutions. The lowest concentration of virus coat protein detected was 1 ng (Plate 4.7.).

Sap was prepared from the 75 regenerated tobacco plants showing kanamycin resistance (section 4.2.5.) and analysed for PVY coat protein. Samples of 35 μl (estimated as 450 μg of total plant protein) were diluted to 200 μl and applied to the NCM in duplicate and probed with primary and secondary antisera. PVY^N infected and uninfected tobacco sap were applied as positive and negative controls. Plate 4.8. shows an example of the blots after colour development. A high degree of non-specific binding was observed as indicated by slots 1 and 2 which contained uninfected tobacco sap samples. No conclusively positive signals were detected from any of the transgenics.

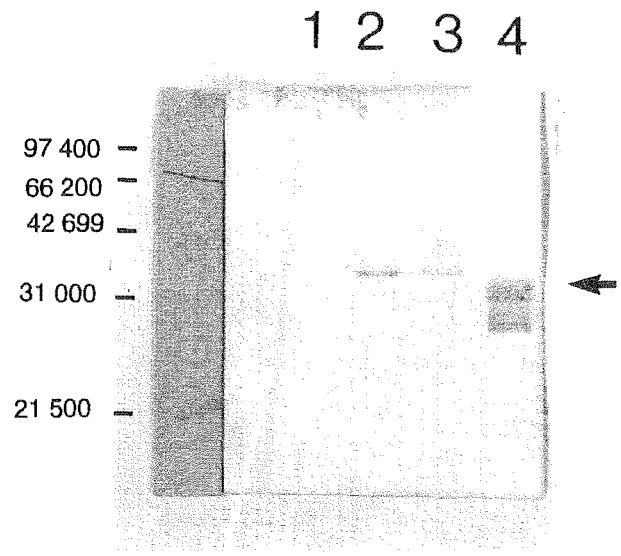


Plate 4.5. Western blot analysis. **Lane 1.** Uninfected tobacco sap. **2.** PVY^{C} infected sap. **3.** PVY^{N} infected sap. **4.** 5 μg purified PSbMV.

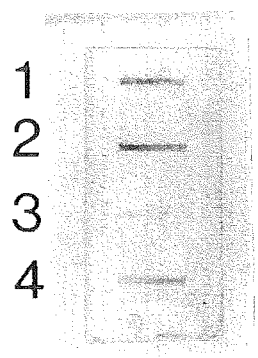


Plate 4.6. Protein slot blot analysis. **Slot 1.** PVY^{N} infected sap. **2.** PVY^{C} infected sap. **3.** Uninfected sap. **4.** 5 μg PSbMV.

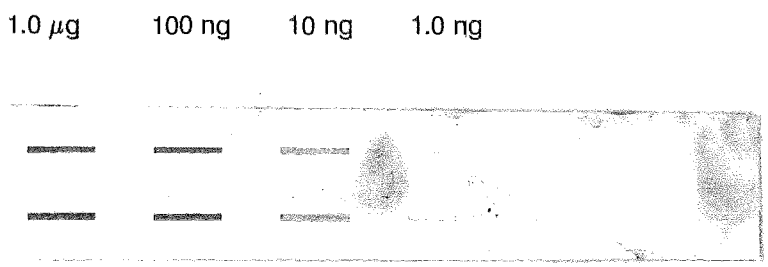


Plate 4.7. Protein slot blot analysis. PVY^{N} (1 mg ml^{-1}) dilution series.

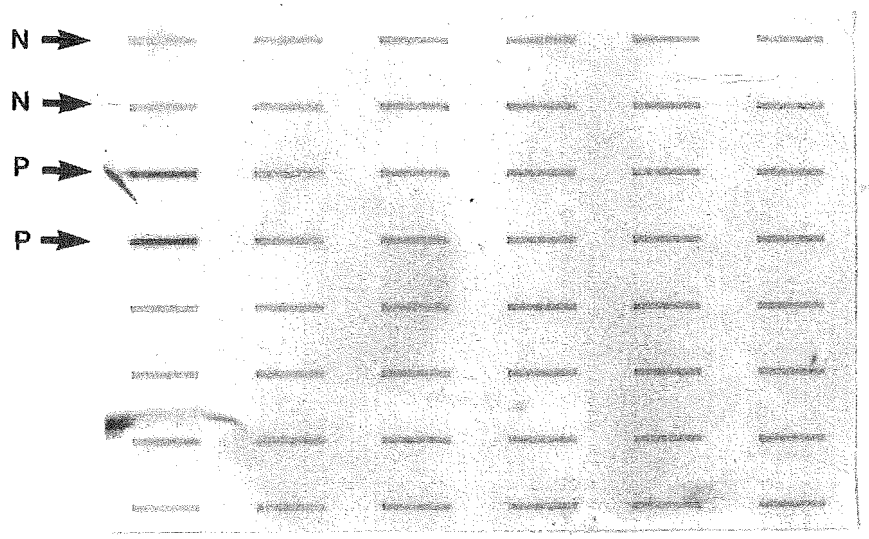


Plate 4.8. Protein slot blot analysis of selected kanamycin resistant *N. plumbaginifolia*. Slots loaded with positive and negative controls are indicated.

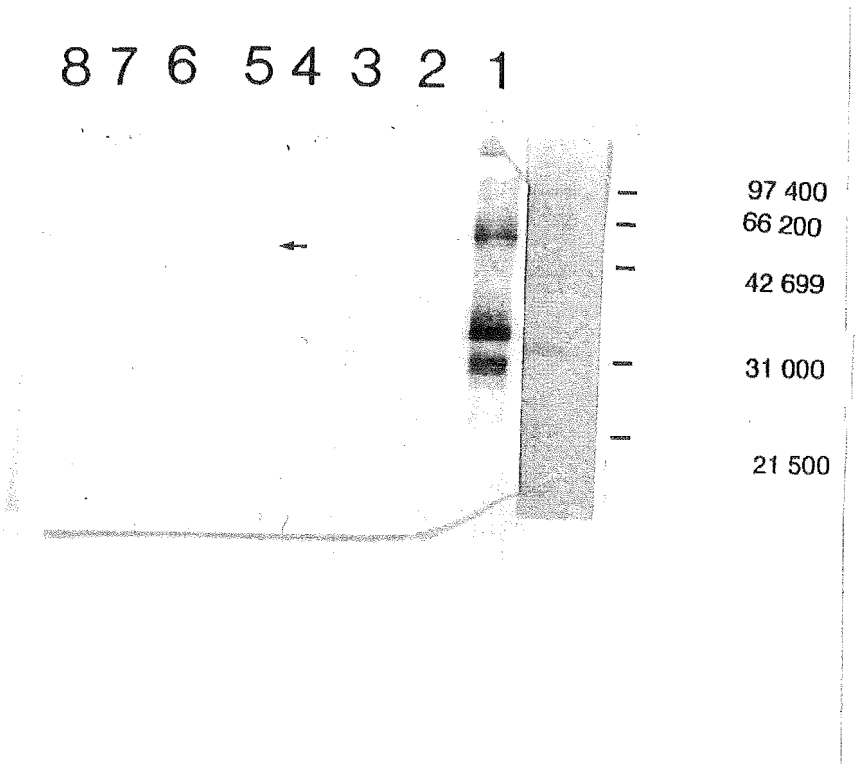


Plate 4.9. Western blot analysis of six kanamycin resistant *N. plumbaginifolia* plants. The lanes were loaded with sap from **1.** PVY^N infected tobacco. **2.** Uninfected tobacco. **3-8.** Plants 3, 5, 6, 28, 29 and 73, respectively.

Six kan^R plants which gave positive signals when probed for the coat protein gene by the DNA slot blot procedure, were further analysed by a Western blot (Plate 4.9.). Samples were prepared and 8 μl samples from each plant (estimated to contain 100 μg of total plant protein) were loaded onto a 12.5% mini-Laemmli gel. A doublet was observed migrating as proteins of 33 kd and 31 kd in the lane loaded with PVY^N infected *N. tabacum*. A faint signal could be detected for plant 6, migrating as a protein with a molecular weight of 56 kd. A similar signal was observed in the lane loaded with PVY^N infected tissue, but in the protein prepared from uninfected tissue. This could be the result of non-specific binding of the antiserum to the large subunit of Rubisco which has a molecular weight of 53 000 and accounts for 50% of the total plant protein (Herrera-Estrella *et al.*, 1984). No detectable protein from the other five kan^R plants was observed to react with the anti-PVY sera. This indicated that PVY^N coat protein was not accumulated by these five plants at levels which could be detected by Western blot analysis.

4.3.5. DNA HYBRIDISATION ANALYSIS OF TRANSFORMED PLANTS

DNA was prepared from 30 kan^R tobacco plants. Control DNA was prepared from uninfected *N. plumbaginifolia* and PVY^N infected *N. tabacum*. Plate 4.10. shows a 1% agarose gel loaded with DNA prepared from 0.8 g of uninfected *N. tabacum* and illustrates the quality of DNA isolated using this procedure.

The DNA was first analysed by slot blot hybridisation. Five to ten μg of prepared DNA were applied to the nylon in duplicate. The membrane was hybridised with 4 $x10^8$ cpm/μg of ^{32}P-labelled probe. Duplicate 1, 3, 6 and 12 pg spots of unlabelled *Cla*I fragment were included as controls. The resulting autoradiograph is depicted in Plate 4.11. Twenty-three plants gave positive signals indicating the presence of the PVY^N coat protein gene in their genomic DNA. These plants were transformed with the gene constructs encoded on pVYN112, pVYN51 and pVYN53 in *Agrobacterium* strain A4T, and pVYN112 in *Agrobacterium* strain LBA4404. DNA prepared from kanamycin resistant plantlets transformed with the gene constructs encoded on pVYN41 and pVYN42 in *Agrobacterium* strain A4T, pVYN113 and pVYN42 in *Agrobacterium* strain LBA4404, and PVY^N infected tobacco sap also gave positive signals. Uninfected tobacco sap and DNA prepared from plants transformed by *Agrobacterium* strain A4T containing the control vector pCGN46 gave no signals.

Six transgenic plants giving strong positive signals for the coat protein gene in the DNA slot blot, were analysed further by the Southern blot procedure. DNA prepared from uninfected *N. plumbaginifolia*, and DNA reconstituted with 6 pgs of unlabelled pVYN27 *Cla*I fragment were included as negative and positive controls. Plate 4.12. shows the stained gel after electrophoresis. The gel was blotted and the nylon probed with 4 $x10^8$ cpm/μg ^{32}P-labelled *Cla*I fragment as described (Plate 4.13.). Positive signals were detected after a three day exposure for plants 3, 6, 28 and 29 at >20 kb, 2180, 1780 and >20 kb base pairs respectively. A second

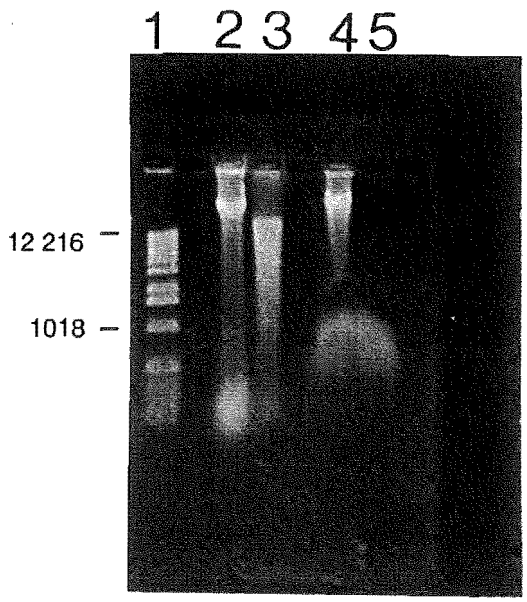
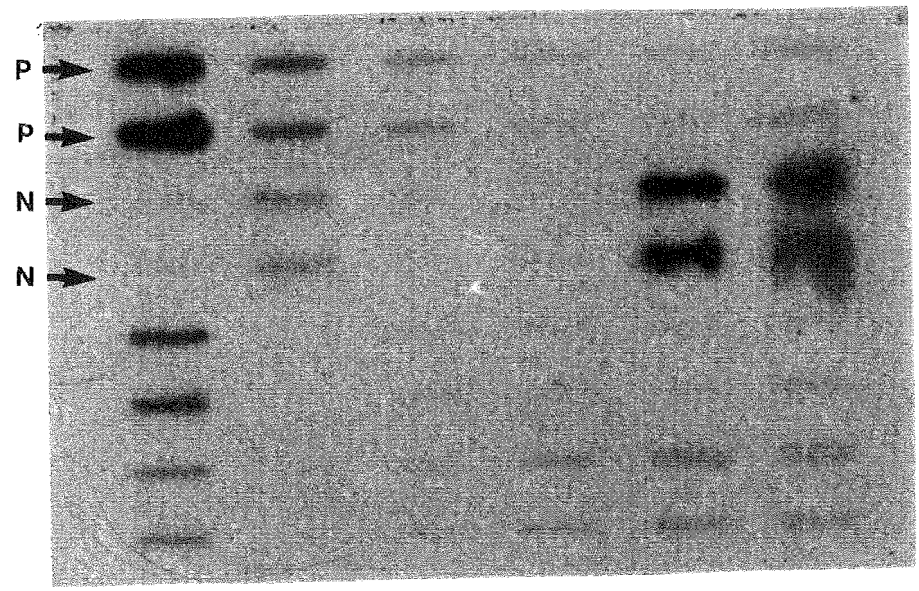
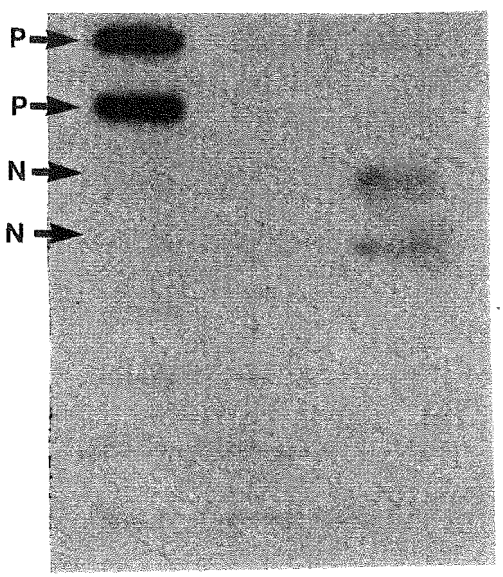


Plate 4.10. 1% agarose gel electrophoresis of DNA prepared from 0.8 g *N. plumbaginifolia* by the method of Dellaporta *et al.* (1983). **Lane 1.** 1 μg BRL 1kb DNA ladder. **2. & 4.** 10 μl undigested DNA. **3.** 10 μl DNA digested for 1 hour at 37^{O}C with 10 units *EcoR*I and 1 μg RNase. **5.** 10 μl DNA treated with 1 μg DNase for 1 hour at 37^{O}C.





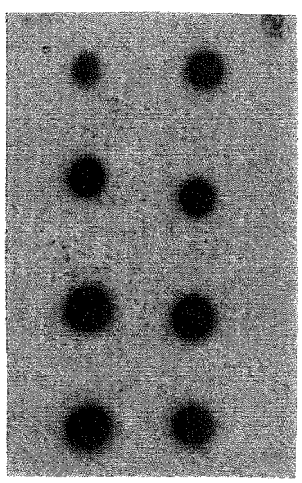


Plate 4.11. DNA slot blot analysis of 30 kanamycin resistant *N. plumbaginifolia* plants. Slots loaded with DNA prepared from PVY^N infected (P) and uninfected (N) plants are indicated.

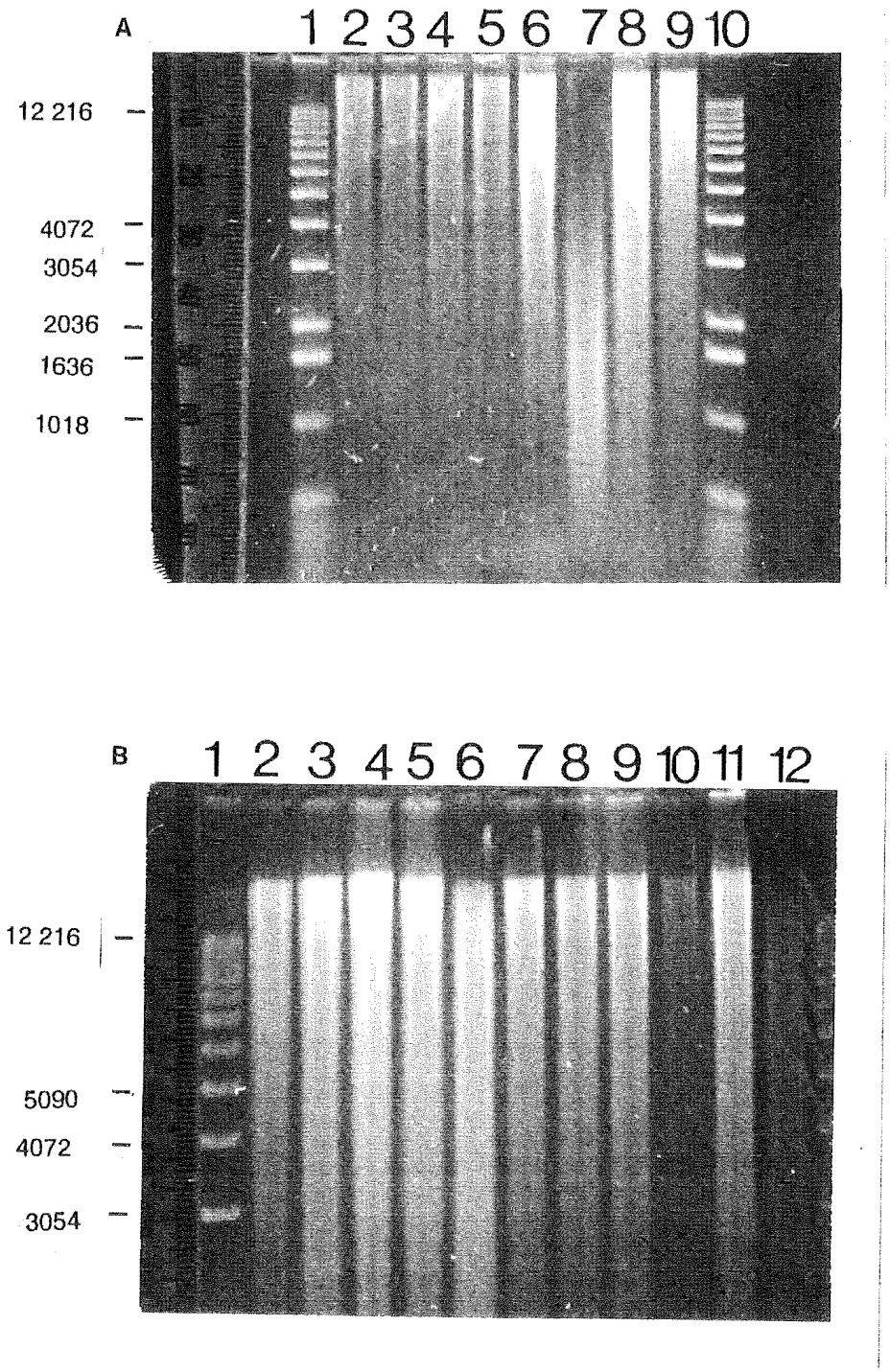


Plate 4.12. 1% agarose gel electrophoresis of *BamH*I digested DNA in preparation for Southern blot analysis. **(A)** Samples to be probed with pVYN27 fragment. **Lanes 1. & 10.** 1 μg BRL 1kb DNA ladder. **2-7.** DNA prepared from plants 3, 5, 6, 28, 29 and 73, repectively. **8.** DNA prepared from an untransformed plant. **9.** Untransformed DNA, reconstituted with unlabelled fragment. **(B)** Samples to be probed with the NPT II fragment. **Lane 1.** 1 μg BRL 1 kb DNA ladder. **2.** Untransformed DNA reconstituted with unlabelled NPT II fragment. **3.** DNA prepared from a PVY^N infected plant. **4.** DNA prepared from uninfected plant. **5-12.** DNA prepared from plants 3, 8, 9, 18, 25, 29, 39 and 73, respectively.

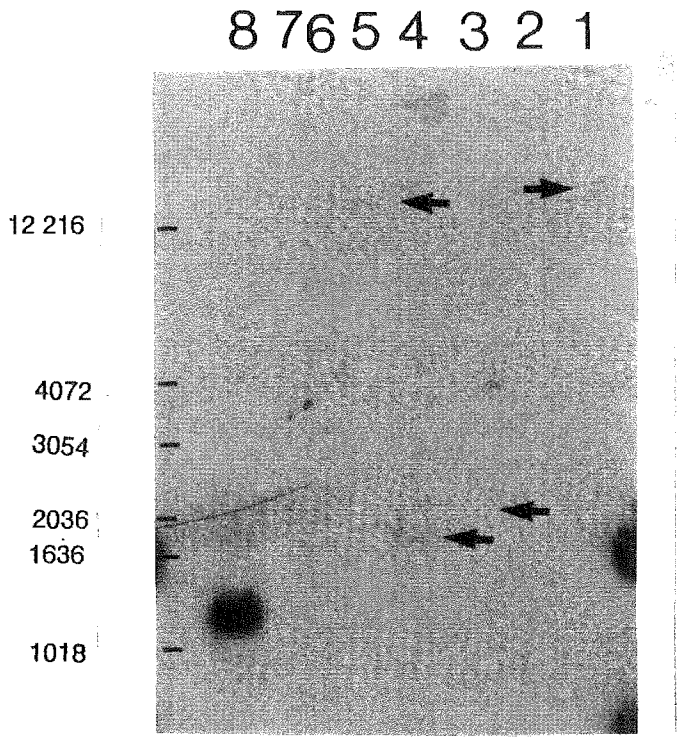


Plate 4.13. Southern blot analysis I. The blot was probed with the ^{32}P-labelled pVYN27 fragment. **Lanes 1-6.** DNA prepared from plants 3, 5, 6, 28, 29 and 73, respectively. **7. & 8.** Negative and positive controls as described Plate 4.12. The positive signals are arrowed.

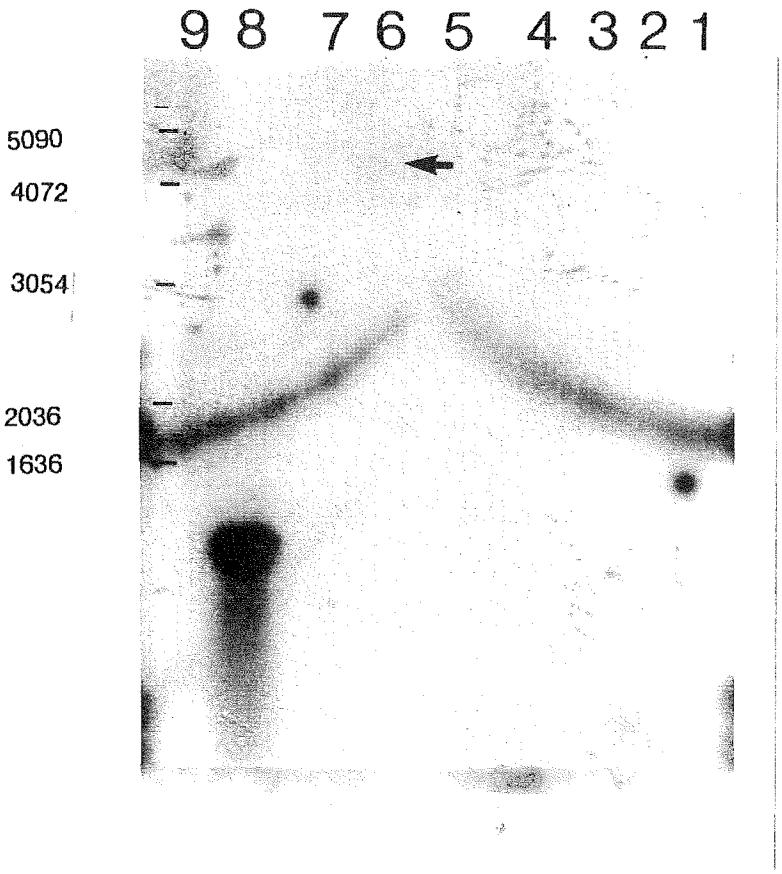


Plate 4.14. Southern blot analysis II. Probed as described for Plate 4.13. **Lanes 1-6.** are loaded as for Plate 4.13. **7.** DNA prepared from an untransformed plant. **8.** Reconstituted DNA. **9.** DNA prepared from a PVY^{N} infected plant. The positive signals are arrowed.

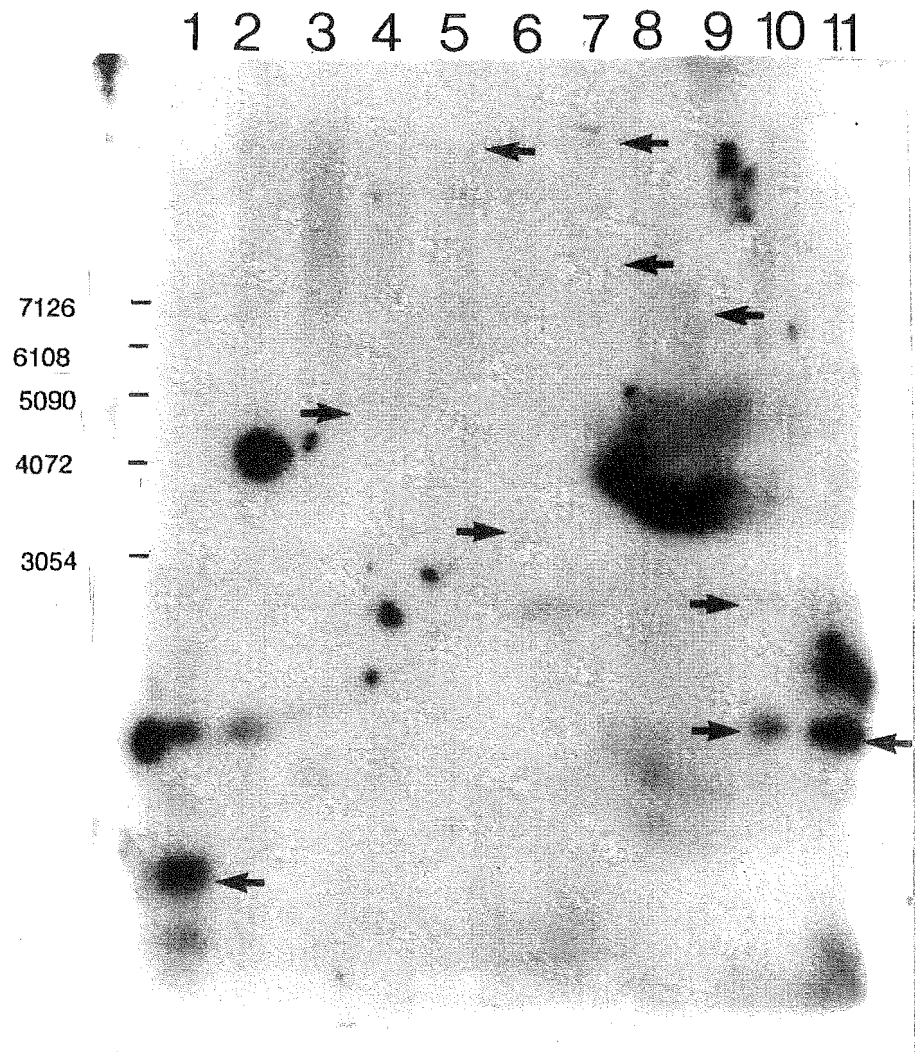


Plate 4.15. Southern blot analysis III. The blot was probed with ^{32}P-labelled NPT II fragment. **Lanes 1-3.** Control DNA (see Plate 4.12 (B)). **4-11.** DNA prepared from plants 3, 8, 9, 18, 25, 29, 39 and 73, respectively. The positive signals are arrowed.

Southern blot produced a signal for plant 73 at 4300 base pairs (Plate 4.14.). Both blots produced a positive signal at approximately 1100 base pairs for lanes loaded with unlabelled *Cla*I fragment. No signal was detected in DNA prepared from plant 5. Plants 3, 6, 28 and 29 were transformed with the chimaeric binary vector pVYN112 and plant 73 with pVYN51. A third Southern blot was probed with 2.8 x10^8 cpm/μg ^{32}P-labelled NPT II fragment. The resulting autoradiograph is illustrated in Plate 4.15. Faint signals were observed in DNA from all the plants analysed. NPT II probe (1.25 kilobase pairs long) hybridised to DNA fragments from plants 3, 8, 9, 29 and 73, which were 4800, >10 000, 3200, 6600 and 1700 nucleotides long respectively. DNA from plants 25 and 39 each produced two bands with fragments >10 000 and >8000, and 2500 and 1700 nucleotides long respectively.

DNA was prepared from kan^R hairy roots which were derived from potato stem explants transformed with the chimaeric and non-chimaeric binary vectors described in section 4.3.2. Three μl samples were applied to a 1% agarose gel and the quality of the prepared DNA was observed (Plate 4.16.). The DNA was then analysed for the PVY^N coat protein gene by slot blot analysis and probed with the 4 x10^8 cpm/μg ^{32}P-labelled *Cla*I fragment. Positive signals were observed for roots transformed with chimaeric binary vectors pVYN112, pVYN113, pVYN51 and pVYN53. No signals were detected for vectors pCGN587, pCGN46, pVYN42 or non-transformed root tissue (Plate 4.17.).

4.3.6. RNA HYBRIDISATION ANALYSIS OF TRANSFORMED PLANTS

The expression of the viral coat protein gene in six transgenic plants was examined at the RNA level by Northern blot analysis. RNA was prepared from the six transgenic plants showing kanamycin resistance which were previously analysed by Southern blots. Five μl samples were electrophoresed on a 1% agarose gel to determine the quality of RNA isolated (Plate 4.18.). Approximately 50 μg of RNA were electrophoresed and analysed by a Northern blot as described. Two faint bands which migrated at 1750 and 2700 base pairs were detected in the RNA from plant 29. A second pair of bands of a similar length (1750 and 2550) were detected in RNA from plant 3. Probe hybridised to a fragment of 1500 base pairs in RNA from plant 6 (Plate 4.19). Plants 3, 6 and 29 were all transformed with pVYN112. A positive signal at approximately 1100 base pairs was detected in the positive control lane, containing unlabelled *Cla*I DNA fragment. Hybridisation signals were not detected in RNA from plants 5, 28 or 73 by the Northern blot, suggesting that the gene inserted in these plants was either not transcribed *in vivo*, or that the transcripts could not be detected.

4.3.7. INHERITANCE OF THE KANAMYCIN RESISTANT PHENOTYPE

Seeds were collected from six kanamycin resistant plants and replicates of 50 seeds were germinated on 1/2 MS in the presence and absence of kanamycin. Germination was defined as

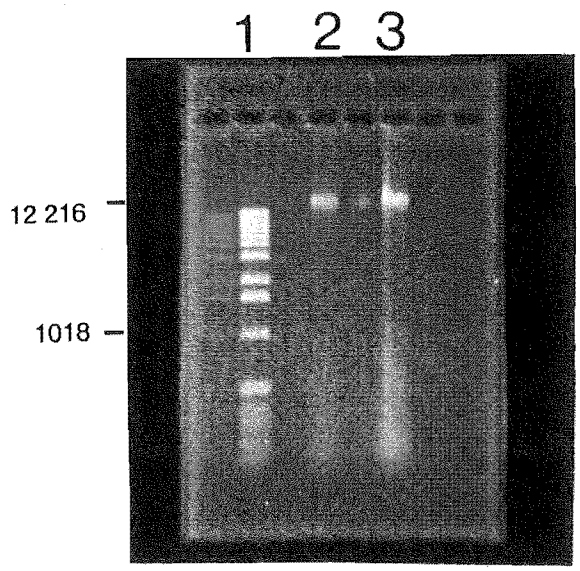
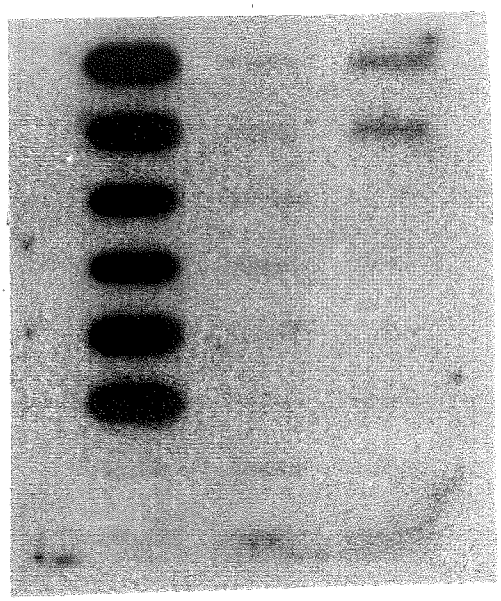


Plate 4.16. Electrophoresis of DNA prepared from kanamycin hairy roots by the method of Dellaporta *et al.* (1983). **Lane 1.** 1 μg BRL 1kb DNA ladder. **2-3.** 5 μl and 10 μl DNA, respectively.

A



B

12 pg unlab *Cla*I frag	pVYN112	pVYN53
"	"	"
6 pg "	pVYN113	pCGN587
"	"	"
12 pg "	pVYN42	pCGN46
"	"	"
non-transf DNA	pVYN51	
"	"	

Plate 4.17. **(A)** DNA slot blot analysis of kanamycin resistant hairy roots. **(B)** Key to samples loaded on the membrane.

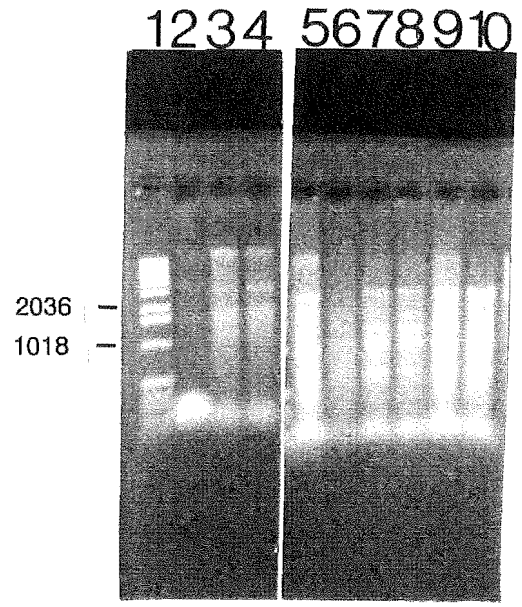


Plate 4.18. RNA prepared from 0.1 g of *N. plumbaginifolia* by the method of Habili *et al.* (1987). **Lane 1.** 1 μg BRL 1kb DNA ladder. **3-8.** RNA prepared from kanamycin resistant *N. plumbaginifolia* plants 3, 5, 6, 28, 29 and 73, respectively. **9. & 10.** RNA prepared from PVY^N infected and uninfected plants.

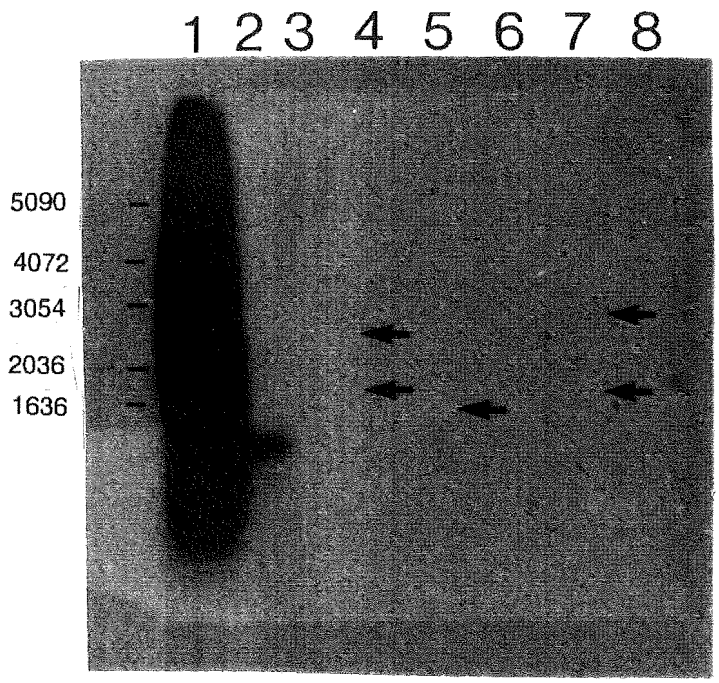


Plate 4.19. Northern blot analysis. The blot was probed with ^{32}P-labelled pVYN27 fragment. **Lane 1.** RNA prepared from a PVY^N infected plant. **2.** RNA prepared from an untransformed plant, reconstituted with 12 pg of unlabelled pVYN27 fragment. **3-8.** RNA prepared from plants 3, 5, 6, 28, 29 and 73, respectively.

the appearance of cotyledons and a primary root. Of the seedlings which germinated on kanamycin containing medium, a number bleached and died after about one week (Table 4.5.).

Table 4.5. Inheritance of kanamycin resistant phenotype.

Plant[a]	+kan. media[b]		Ratio[c]	Total germination[d]	
	bleached	green	bleached:green	+kan	-kan
3	17	67	1:3.9	84	82
5	12	54	1:4.5	66	76
6	11	58	1:5.3	69	88
9	10	60	1:6	70	86
28	9	65	1:7.2	74	66
29	3	19	1:6.3	22	56
73	16	55	1:3.4	71	70
untransformed (*in vitro*)	89	1	-	90	84
untransformed (commercial)	69	0	-	69	76

a. Selected plants showing kan^R.

b. Data collected from germinating a total of 100 seeds (duplicates of 50) on media containing kan. (400 mg L^{-1}). Seedlings either bleached and died (bleached) or remained healthy (green).

c. Data collected to test for the Mendelian inheritance of the NPTII gene.

d. Total number of seeds germinating on media with or without kan. (400 mg L^{-1}). Seeds not germinating were considered 'non-viable'. '+kan' includes 'bleached' and 'green' seedlings.

Almost all the seedlings from untransformed sources bleached after germination, showing that progeny from transformed plants were resistance to kanamycin. This indicated that the NPT II gene was inherited by the progeny of the transformed plants.

Table 4.6. Chi square goodness of fit test for homogeneity between transformed plants in terms of percentage germination.

Plant	Germinating	Non-viable	
3	84	16	***
5	66	34	
6	69	31	
9	70	30	
28	74	26	
29	22	78	***
73	71	29	

A Chi square test for homogeneity in terms of germination percentages between transformed plants was performed (Table 4.6.). Plants 3 and 29 showed a significant difference ($p<0.001$) from the expected mean percentage germination of 65%. Examination of the raw data showed plant 3 to have a higher percentage germination compared to the other transgenic plants. Plant 29 showed a lower number of seeds germinating and consequently, a higher number of non-viable seed. It can be assumed that this was peculiar to the individual and that the NPT II gene did not alter the initial germination pattern by conferring a higher number of non-viable seed to the parent plant.

The difference in terms of germination between non-transformed plants and the transgenic plants as a group was tested by a Chi Square test of association. The expected value of 80% germination for untransformed plants and the observed value of 75% for transformed plants was shown to be not significant.

Table 4.7. Chi square goodness of fit test for 3:1 ratio between expressors and non-expressors of the NPTII gene.

	Number of F1 seedlings		
Plant	Expressing NPTII	Not expressing NPTII	χ^2 (3:1)
3	67	17	1.00
5	54	12	1.64
6	58	11	3.01
9	60	10	4.27 *
28	65	9	6.5 *
29	19	3	1.58
73	55	16	0.23

If transgenic plants express the coat protein at a single locus, 75% of seedlings would be expected to contain the NPT II gene (Mendelian inheritance). Seedling progeny of the six kan^R plants analysed by Southern and Northern blots were examined for a 3:1 ratio based on green:bleached data, by a Chi Square test goodness of fit test (Table 4.7.). Plants 3, 5, 6, 29 and 73 showed evidence of a Mendelian segregation. The segregation ratio of the seedlings from plants 9 and 28 did not fit the 3:1 expression model ($p<0.05$). This suggests the coat protein gene was expressed from genes inserted at multiple loci. Similar observations were made by Powell Abel *et al.* (1986) for the progeny of tobacco transformed with the TMV coat protein gene.

4.3.8. CROSS PROTECTION EXPERIMENT

Two weeks after the inoculation of selected plants with the PVY^N inoculum, the growth room overheated and the plants were all severely damaged. Prior to this, no symptoms had been observed on any of the plants including the non-transgenic control plants.

4.4. DISCUSSION

4.4.1. PLANT TRANSFORMATION

Nicotiana plumbaginifolia and *Solanum tuberosum* cv. Iwa plants were transformed with the PVY^N coat protein gene for the following reasons. Tobacco is a particularly suitable 'model' plant for transformation studies as it is amenable to tissue culture techniques. Potato, on the other hand, is agronomically important but is less amenable to tissue culture techniques. Also, potato plants regenerated from both protoplasts and explants have shown considerable phenotypic and genotypic variation (Ooms *et al.*, 1985). In this study, potato leaf disks were transformed using *Agrobacterium* suspensions from both LBA4404 and A4T strains. Clumps of callus developed from single kanamycin resistant colonies, but no regeneration of shoots was achieved. Other workers have observed that the leaf disk method was not applicable to all plant species and has proved particularly difficult to adapt for potato (Shahin and Simpson, 1986; Draper *et al.*, 1988). The transformation of potato stem sections with A4T (Ooms *et al.*, 1983, 1985) was attempted as an alternative method for introducing the viral gene.

Tobacco leaf disks were transformed with the A4T strain of *Agrobacterium*, and many discrete kanamycin resistant callus colonies were observed. However, the transformation and regeneration time for colonies infected with the LBA4404 strain of *Agrobacterium* was longer (by weeks) and fewer transformants were obtained compared with the A4T strain. Armitage *et al.* (1988) suggested that the difficulty in obtaining transformants with binary vectors usually reflects the limitations of the pALA4404 virulence functions. This is interesting as the *Agrobacterium* strain, LBA4404 is commonly used for plant transformations (van Dun *et al.*, 1987; Bevan *et al.*, 1985; Baulcombe *et al.*, 1986; Hoekema *et al.*, 1983). On appropriate media, potato stem explants infected with the *Agrobacterium* strain A4T, were successfully redirected from hairy root proliferation, to the production of callus and the formation of shoot initials from single transformed colonies. The exploitation of different *Agrobacterium* strains may possibly facilitate faster and more efficient transformation of *Solanaeceous* hosts, and may aid in the transformation of other dicotyledonous species.

4.4.2. TESTS FOR THE TRANSGENIC NATURE OF THE TRANSFORMANTS

To confirm the transfer and integration of T-DNA carrying the PVY^N coat protein gene into tobacco and potato chromosomes, genomic DNA was prepared from leaf tissue and hairy roots and analysed by DNA slot blot hybridistion. A positive hybridisation signal was detected in 23 of the 30 kanamycin resistant tobacco plants tested. The majority of these plants had been transformed with the chimaeric binary vector pVYN112, which contained the coat protein transcriptional fusion construct. Others were transformed with pVYN53, pVYN51, pVYN42 and

pVYN41 (Table 4.3.). The seven plants giving no hybridisation signals were all transformed with the pVYN112 binary vector. DNA from hairy roots was also probed for the PVY^N coat protein gene. Positively hybridising sequences were observed for roots containing the expression constructs and the 'sense' and 'anti-sense' sequences under the direction of both the 35S and mannopine synthase promoters.

Jordon and M^{c}Hughen (1988) stated that the transgenic nature of transformed plants must be proven by Southern blot analysis or by a demonstration of the inheritance of the genetic trait. They also suggested rooting in selective media to be a good criterion for the selection of true transgenics. Draper *et al.* (1988) state that because roots are more sensitive to antibiotics than shoots, the ability to produce roots on selective media is a good indication of the transformed nature of the plant. The presence of the kanamycin resistant phenotype (conferred by the NPT II gene and encoded by the binary vector) would be expected to correspond with the presence of the chimaeric PVY^N gene. Both hairy roots derived from transformed potato stem explants, and regenerating plantlets derived from transformed tobacco leaf disks, were maintained on media containing kanamycin at concentrations normally inhibitory to plant growth. Tobacco shoots were rooted in medium containing 100 μg ml^{-1} kanamycin before being transplanted into potting mix. The stringency of selection was monitored by placing untransformed plants at different developmental stages, on kan^+ media.

Analysis of the seedling progeny of transformed tobacco plants was undertaken. Self-fertilised seed from six plants which gave positive hybridisation signals for the PVY^N coat protein gene, were germinated on kan^+ media and the data analysed by a Chi square test for goodness of fit for a 3:1 ratio. The seedling progeny of four of the six plants segregated with a ratio consistent with the expression of the gene from a single locus. The seedling progeny from the two other plants appeared to express the NPT II gene from more than one locus.

Proof of the transgenic nature of regenerated plants comes from the presence of the chimaeric gene as demonstrated by DNA hybridisation. Expression was confirmed by callus production (and hairy roots proliferation) on media containing kanamycin at normally inhibitory concentrations. The regenerated tobacco plants were morphologically normal and produced viable seed. The progeny derived from several transformed plants were shown to inherit the kanamycin phenotype (and by association, it is expected the PVY^N coat protein gene) in a Mendelian manner.

4.4.3. NORTHERN AND SOUTHERN BLOTTING

Using *N. plumbaginifolia* as a model host system unequivocal evidence for the presence of the PVY^N coat protein gene and the NPT II gene was sought by Southern blot hybridisation. DNA prepared from six transgenic plants was analysed according to Southern (1975). The expected size of the *BamH*I fragments containing the appropriate gene and hybridising to both probes was >12.1 kilobase pairs. DNA from five plants gave positive hybridisation signals when probed with

the PVY^N gene, with two signals (from plants 3 and 29) being >12.1 kb long. DNA from all seven plants tested gave positive signals when probed with the NPT II gene sequence. Of these, two had fragments hybridising which were >12.1 kb long. Control lanes loaded with unlabelled probe fragments gave positive signals at the appropriate electrophoretic mobilities corresponding to 956 bp for PVY^N and 1.25 kb for NPT II. Signals <12.1 kb were detected and cannot be explained, except as due to the incomplete transfer of T-DNA from *Agrobacterium* to the plant chromosome.

The transcriptional activity of the chimaeric PVY^N gene in these six cloned tobacco plants was analysed by Northern blot hybridisation. RNA blots were probed with the 956 bp gene sequence encoding the PVY^N coding region. The approximate size of the chimaeric mRNA was expected to equal the sum of the cDNA (956 nucleotides) plus the 5'-untranslated sequence for CaMV 35S promoter (111 nucleotides) and 190 nucleotides contributed by the 3'-*tml* sequence, totalling 1257 nucleotides, excluding the poly(A) tail. A 1500 nucleotide transcript was detected in RNA from plant 6. 1750 and approximately 2500 long transcripts were detected in the RNA from plants 3 and 29. It was noted that bands corresponding to the reconstituted control DNA were migrating as slightly longer fragments than expected. This indicated that the 1500 and 1750 bp fragments detected could be mRNA transcripts from the PVY^N coat protein gene. The 2500 bp transcripts could either be due to aberrant initiation upstream of the CAP site, or to aberrant termination downstream of the polyadenylation signal. Alternatively, the signals could be from 18S and 23S ribosomal mRNA transcripts (1926 and 3580 nucleotides respectively). However, these bands were not detected in the RNA from the other plants and similar amounts of RNA were loaded in all lanes.

Both the DNA and RNA purification protocols yielded high molecular weight nucleic acid, and digestion of the DNA with *BamH*I restriction endonuclease was complete (Plate 4.12.). The specific activity of the probes were at recommended levels for hybridisation (Scott *et al.*, 1988). It is presumed that these inconclusive results for Southern and Northern blotting are due to an inefficient transfer of long DNA and RNA species to the nylon membrane. Scott *et al.* (1988) suggest that longer DNA species are difficult to transfer, and this observation was repeated by G. Timmerman (pers. comm.). Both sources suggest that acidic depurination of DNA followed by alkaline blotting gives a more efficient transfer of long DNA. As DNA fragments >12.1 kilobase pairs were expected to be detected by Southern blots from transformed plant DNA, inefficient transfer may explain their absence. An additional factor contributing to the poor quality of blots obtained may have been the low stringency of the hybridisation washes. Due to the inadequate resources and finances available for this study, the conditions for transfer could not be optimized and the experiments could not be repeated.

Southern and Northern blots indicated the presence of the PVY^N coat protein and NPT II genes, and that the PVY^N coat protein is being transcribed *in vivo*. They provide evidence which is supportive, but not unequivocal, for the transgenic nature of the regenerated plants.

4.4.4. PVY COAT PROTEIN EXPRESSION

Seventy five kanamycin resistant, regenerated tobacco plants were examined for coat protein accumulation by protein slot blot analysis using antiserum to the PVY coat protein. Six plants were further examined by the Western blot procedure. Scott *et al.* (1988) suggest that 50 ng of a selected antigen can be detected by the Western blot procedure. The Biorad blotting manual reports the detection of 10-50 pg of antigen by Western blots using a secondary antibody conjugated to alkaline phosphatase. The protein slot blot experiments reported in this study show that 1 ng of purified coat protein could be detected using a primary antibody and a secondary antibody conjugated to alkaline phosphatase. Ten ng of PVY^N was detected in crudely extracted total protein in a reconstitution experiment (data not shown). This demonstrates the sensitivity of these assays and implies that if the virus coat protein was expressed *in vivo*, it could be detected. Bevan *et al.* (1985) reported TMV coat protein accumulating to a level of approximately 0.002% of the total soluble cell protein. Powell Abel *et al.* (1986) observed that eight different transgenic plants produced TMV coat protein to 0.1% of the extractable cellular protein. Nelson *et al.* (1987) observed 0.05% - 0.1% TMV coat protein accumulation, and the expression of PVX coat protein detected by Western blots was approximately 0.02% - 0.1% of total protein (Hemenway *et al.*, 1988). If PVY^N coat protein accumulated to a level similar to these reported for other viruses, an estimated 450 ng and 100 ng of coat protein was loaded during slot blot and Western blot hybridisations, respectively. These quantities of protein should easily be detected given the sensitivity of the assays used. It therefore appears that the PVY^N coat protein gene was not expressed at a detectable level in any of the transgenic plants tested.

There are a number of explanations for this lack of expression. Possibly the complete coat protein gene was not cloned, or the constructs were not optimally engineered for expression. Due to inconclusive data from Northern hybridisations, it is not clear whether the chimaeric gene was being transcribed, or whether the transcripts were being recognised and translated into coat protein by the host. Insertion of the PVY^N coat protein cDNA into a plasmid in *E. coli* for *in vitro* transcription and translation may help define which process is obstructed *in vivo*. Comparison of the PVY^N coat protein sequence data with published potyviral sequences (Chapter 3.0) does not reveal any obvious deletions or insertions in the sequence and implies that the gene is complete. The expression cassettes were constructed specifically for translation in a eukaryote host and sequences for ribosome binding, translation initiation and termination have all been identified. The process of T-DNA integration, as far as is known, is a random event. The position of the gene in the plant genome could account for the lack of expression.

Powell Abel *et al.* (1986) and Bevan *et al.* (1985) observed apparent differences in the levels of expression for a similar gene (TMV coat protein) in transgenic tobacco plants. Powell Abel *et al.* attribute this to differences in the CaMV 35S promoter fragment, and to the host plant genotypes used. Both arguments could be valid in the present study. The expression cassettes

constructed for the PVY^N coat protein contained a minimally trimmed 35S promoter and produced transcripts with a long leader sequence (approximately 111 nucleotides). Powell Abel *et al.* (1986), van Dun *et al.* (1987) and Nelson *et al.* (1987) used truncated 35S promoters with 42, 20 and 58 long nucleotide leader transcripts respectively. Bevan *et al.* (1985), however, used a construct with a longer 35S 5'-leader sequence of 132 nucleotides. They observed expression and coat protein accumulation to only 0.002% of the total soluble protein. It is unlikely that accumulation of PVY^N coat protein at this level would be detected by protein slot blots or Western blots, given the sensitivity of the assays. Yamaya *et al.* (1988) described the expression of TMV RNA in transgenic tobacco plants using a modified 35S promoter. They suggested that for the RNAs produced *in vitro* to be infectious the 5'-end of the *in vitro* transcripts must co-incide with the 5' end of the native viral RNAs. They noted that the addition of six extra nucleotides at the 5'-end of TMV cDNA reduced the infectivity of the transcripts approximately 100 fold. Possibly a longer CaMV 35S 5'-untranslated leader sequence may destabilise the mRNA species. Whatever the mechanism, the studies of Powell Abel *et al.* (1986), Nelson *et al.* (1987), van Dun *et al.* (1987), Bevan *et al.* (1985) and Yamaya *et al.* (1988) seem to suggest a correlation between long leader sequences and unexpressed transcripts. R. Beachy (pers. comm.) endorsed the shortening of the 5'-leader to increase expression of the coat protein. Bevan *et al.* (1985) suggest that the differences in the stability of chimaeric mRNAs may explain the phenomenom of different steady-state levels of CaMV 35S transcripts. Gallie *et al.* (1989) confirm that the requirements for efficient scanning for the AUG start codon by the ribosome is thought to be a short leader with no secondary structure.

Another factor which may explain the absence of gene expression is presented by Hepburn *et al.* (1983). They reported that all T-DNA encoded nopaline synthase gene copies were methylated to some extent and that this correlated with an extremely low level of gene expression. Subsequent demethylation is paralleled by an increase in gene transcription. Similar observations have been made in animal cell lines (Hepburn *et al.*, 1983).

R. Beachy (pers. comm.) commented that many plants must be screened in order to find plants protected against potyviral infection. Possibly insufficient plants were tested for expression in this study. Beachy tentatively suggested that the C residue at position +5 is important for expression, and that negative effects can be caused by the poly(A) signal. Gallie *et al.* (1989) suggest that in plants the first 25 poly-adenine residues are the most important for the expression of the corresponding gene, and that a tail of less than 15 residues severely limits the expression of genes in plants.

4.4.5. GENE CHIMAERIC CONSTRUCTS

A number of chimaeric vectors (Table 4.3.) were constructed with a long term view toward: 1. performing cross protection experiments, and 2. testing two different promoters for levels of expression in the host plant system. Two expression vectors were constructed containing the

cDNA of the coat protein of PVY^N and either the 35S promoter from CaMV and a polyadenylation signal from an *Agrobacterium* encoded tumour inducing gene (*tml*), or the *Agrobacterium* encoded mannopine synthase promoter and a polyadenylation signal from the octopine synthase gene.

Agrobacterium-mediated transformation of plant species has created a demand for a greater understanding of the variables controlling the expression of re-introduced genes. Genes require regulatory sequences for expression in transformed plants (Harpster *et al.*, 1988). In a comparitive study, Harpster *et al.* observed that the CaMV 35S promoter provided the highest levels of gene expression in callus tissue produced by a number of transformed explants. Lower expression levels were observed from the constitutive *Agrobacterium* T-DNA promoters 2' and 1', and the nopaline synthase promoter, in that order. Most chimaeric genes made for plant transformations use the 35S promoter (Powell Abel *et al.*, 1986; van Dun *et al.*, 1987, 1988a, 1988b; van Dun and Bol, 1988; Yamaya *et al.*, 1988; Bevan *et al.*, 1985; Hemenway *et al.*, 1988; Baulcombe *et al.*, 1986). In this study, it was hoped that the expression of genes directed by a fourth promoter, the mannopine synthase promoter, could be compared with those from the CaMV 35S promoter. Due to inadequate results from Northern blots, no quantitative comparison of the relative strengths of the two promoters could be concluded. No accumulation of coat protein was detected from transcripts from either promoter.

Palukaitis and Zaitlin (1984) proposed a role for antisense RNA in the mechanism for cross protection. Hemenway *et al.*, (1988) reported that transgenic plants engineered to express an antisense potato virus X (PVX) coat protein transcript were protected against infection at low inoculum concentrations. Other workers (J. Bol, pers. comm.) observed that antisense RNA afforded no protection against infection by alfalfa mosaic virus. Chimaeric vectors with the PVY^N coat protein gene oriented in either a sense or antisense direction, were constructed to test the mechanism for cross protection. A transcriptional fusion designed to express the PVY^N coat protein gene optimally *in vivo* was also constructed. It contained the PVY^N coat protein gene ligated to a synthetic oligonucleotide, and was inserted between 35S promoter and *tml* 3'-sequences. Sequencing data (Chapter 3.0) indicated that the coat protein gene was missing three amino-terminal amino acids (alanine, asparagine and aspartic acid). The synthetic oligonucleotide was designed to encode these in addition to sequences coding for the two amino acids removed by the ligation event, a translation start codon (AUG) and *Xba*I and *Cla*I recognition sites. A sequence postulated to be optimal for eukaryote ribosomal binding (Kozak, 1986) was also encoded. As a result, 10 nucleotides were added to the 35S leader transcript. Following transformation into a binary vector, constructs were mobilised into each of two strains of *Agrobacterium*.

A preliminary cross protection experiment was initiated to determine whether transgenic *N. plumbaginifolia* plants could suppress symptoms caused by PVY^N infection. No symptoms were observed on the inoculated plants, including the non-transgenic controls prior to the growth

room overheating. Possibly the inoculum was too low to cause a systemic infection within the two week period. However, comparable concentrations of virus inoculum have been used in similar experiments to successfully initiate detectable symptoms (Nelson *et al.*, 1987; Powell Abel *et al.*, 1986; van Dun *et al.*, 1987). Unfortunately this experiment could not be monitored further due to the death of the plants.

The use of parent transgenic plants constituted a limiting factor in this experiment as they failed to give a sufficient number of uniform plants to carry out a statistically valid experiment. A more meaningful cross protection experiment requires seedling progeny from self-fertilised transgenic plants to be used. Seed has been collected from transformants shown by DNA hybridisation assays to contain the PVY^N gene. These, along with other transformed tobacco plantlets have been transferred to CRD, DSIR, Private Bag, Christchurch for further regeneration and subsequent genetic analysis.

CHAPTER FIVE

THESIS REVIEW

Two strains of Potato Virus Y (PVY), a local necrotic strain PVY^N and a PVY^C strain, were purified from laboratory host plants, *Nicotiana tobacum* cv. White Burley by the method of Reddick and Barnett (1983). Potyviruses are prone to aggregation and occur in low concentrations in infected cell sap. Features of the purification protocol which overcame these problems included PEG precipitations to concentrate the virus, resuspension of virus pellets in the presence of 1% Triton X-100 and isopycnic centrifugation in a density dependent Cs_2SO_4 gradient. Yields of virus were between 1-5 mg kg^{-1} of infected leaf material. Yields of PVY^N were consistently higher than those of PVY^C.

RNA was isolated from dissociated virus particles and fractionated on a 10%-30% sucrose gradient according to Brakke and van Pelt (1970). Typically, 5 mg or more of freshly purified virus were required to isolate detectable RNA. Full-length RNA species were pooled and used as a template to produce oligo(dT) primed cDNA which was subsequently made double stranded using a modified DNA polymerase I/RNase H protocol (d'Alessio *et al.*, 1987). *Xba*I linkers were ligated onto the cDNA and the double stranded cDNA was inserted into the plasmid vector, pUC19. The resultant colonies were screened and two of the recombinant colonies derived from PVY^C sequences were characterised further. pVYC11 was shown to contain viral 3'-cDNA sequences by DNA hybridisation, while pVYC5 did not. DNA sequencing confirmed that pVYC5 contained no viral sequences and it was concluded that the recombinant plasmid contained host plant gene sequences.

A recombinant clone containing PVY^N gene sequences was characterised by DNA sequencing (Sanger *et al.*, 1977). The clone encoded the 3'-terminal 1134 nucleotides which contained one large open reading frame capable of encoding a protein of 264 amino acid residues with a combined molecular weight of 29 631. Comparison of this amino acid sequence with the published coat protein sequences for two other PVY strains (Shukla *et al.*, 1986; van der Vlugt *et al.*, 1988) confirmed that the 3'-cistron coded for the viral coat protein of PVY^N. The ten amino-terminal amino acids were confirmed by modified Edman degradation. A 326 nucleotide, untranslated sequence terminating in a polyadenylate tract was observed adjacent to the 3'-end of the coding region. An alignment of the PVY^N amino acid sequence with coat protein data from six other potyviruses revealed significant sequence homology in the internal and carboxy-terminal regions. Only PVY^N, PVY^D and PeMV showed discernable sequence similarity in the amino-terminal region. The hydropathy profiles of the seven potyviruses confirmed these regions of sequence homology. The significant amino acid sequence similarity between PVY^N, PVY^D and PeMV (91%-93%) suggested that PeMV should be regarded as a PVY strain and not a

distinct virus as previously thought. An analysis of the 3'-untranslated nucleotide sequence of six potyviruses revealed PVY^N and PeMV to be the only viruses displaying homology in this region. Further analysis of the 3'-untranslated sequence of PVY^N revealed four regions of sequence similarity, two potential regions of secondary structure and three pairs of directly repeated sequences. These were compared to similar features found in PeMV. The codon usage profile for the PVY^N coat protein gene resembled that of unicellular organisms with a preference for adenine and thymidine residues in codon position III (Murray *et al.*, 1989). The PVY^N coat protein was analysed for secondary structure (Garnier, 1978) and revealed ten regions of α-helix, four regions of β-sheet and fourteen regions of coil and/or turn. Two unrooted, equally parsimonious phylogenetic trees were predicted for the seven potyviruses based on coat protein sequence data (Felsenstein, 1987).

Chimaeric genes were constucted to encode the cDNA coding region of the coat protein gene, and either the mannopine synthase promoter and octopine synthase 3'-polyadenylation sequences, or the CaMV 35S promoter and the *tml* 3'-polyadenylation sequences. These chimaeric genes were inserted between the left and right borders of an *Agrobacterium* binary vector encoding the NPT II gene for kanamycin resistance, and mobilised into the *Agrobacterium* species, *A. tumefaciens* (LBA4404) and *A. rhizogenes* (A4T) by a tri-parental mating. *Solanum tuberosum* stem explants and *Nicotiana plumbaginifolia* leaf disks were transformed with the *Agrobacterium* strains containing the tumour- and hairy root-inducing plasmids and the chimaeric gene for NPT II and PVY^N.

Hairy roots regenerating from *S. tuberosum* stem explants transformed with the *Agrobacterium* strain A4T containing the chimaeric gene grew on media containing kanamycin at normally inhibitory concentrations. The presence of the PVY^N gene was confirmed by DNA slot blot hybridisation analysis. Regeneration of the hairy roots into potato plants was not completed. Tobacco plants regenerated from *N. plumbaginifolia* leaf disks transformed with the *Agrobacterium* strains LBA4404 and A4T containing the chimaeric gene, were morphologically normal and fertile. The presence of the chimaeric PVY^N coat protein gene was demonstrated by DNA slot blot hybridisation analysis, and its expression in plant tissues was confirmed by the ability of leaf segments to callus on media containing kanamycin at concentrations that were normally inhibitory. The progeny derived from several transformed tobacco plants were kanamycin resistant and inherited the foreign gene in a Mendelian manner. Preliminary Southern blots on selected transgenic tobacco plants suggested that single copies of the PVY^N and NPT II genes were inserted. Preliminary Northern blots indicated that three of the six tobacco plants investigated were transcribing mRNA from the coat protein gene. No accumulation of coat protein was detected in any of the transformants either by Western blot analysis or by protein slot blots, suggesting that the PVY^N coat protein gene was not expressed at a detectable level in tobacco.

It is hoped that the continued regeneration of whole transformed tobacco and potato plants at CRD, DSIR, Private Bag, Christchurch, will allow a controlled and statistically valid cross protection experiment to be undertaken. Further experiments to compare the genomes of PVY^{C} and PVY^{N} strains may be interesting because although they are closely related, the symptoms they produce in the same host plants are markedly different. Such experiments may facilitate a greater understanding of host-virus interactions about which little is currently known.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr Gail Timmerman for her supervision, for reading draft copies of my thesis and for technical assistance with large scale plasmid preparations and sequencing. I am grateful to John Fletcher of the Plant Diseases Division, DSIR for his advice and help especially in the initial stages of my research; also to Dr Dharma Shukla for advice and discussion, and the many copies of unpublished manuscripts he gave me. Many thanks to David Burritt for his assistance with tissue culture techniques and for his photographic skills. I would especially like to thank Dr Dianne Hill and Gillian Hughes from the Department of Biochemistry, University of Otago for their patient teaching of DNA sequencing and for a great confidence boost. *Agrobacterium* strains, plasmid vectors and sources of *in vitro* tobacco and potato plants were gratefully recieved from Dr Tony Conner, Crop Research Division, DSIR. I thank Dr Peter Shultz and members of the virology lab. at Christchurch Hospital for the use of their ultracentrifuge. Assistance is acknowledged from Manfred Ingelfield for transmission microscopy, Dr Chris Frampton for statistics advice and Jane Chamberlain for assistance with word processing.

I would like to express my special thanks to Vicky Calder for her support, friendship and collaboration during the time of my study. My thanks to her also for reading the draft copies of my thesis. To my family, my heartfelt thanks for your love and support. Finally, my friends, flatmates and work colleagues are thanked most sincerely for their friendship and help. A special ackowledgement to Greg Offer for his time and tolerance.

Financial support was provided by a University Grants Committee Post-graduate Scholarship. Further financial assistance was provided by the Georgetti Scholarship Board, the Winifred Eliza May Scholarship Trust, and by the Crop Research Division, DSIR, Christchurch.

REFERENCES

Abu-Samah, N., and Randles, J. W. 1981. A comparison of the nucleotide sequence homologies of three isolates of bean yellow mosaic virus and their relationship to other potyviruses. *Virology* **110**:436-444.

Ahmed, A. 1987. The use of transposon-promoted deletions in DNA sequence analysis. *Methods in Enzymology* **155**:177-204.

Allison, R., Johnston, R. E., and Dougherty, W. G. 1986. The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology* **154**:9-20.

Allison, R. F., Dougherty, W. G., Parks, T. D., Willis, L., Johnston, R. E., Kelly, M., and Armstrong, F. B. 1985a. Biochemical analysis of the capsid protein gene and capsid protein of tobacco etch virus: N-terminal amino acids are located on the virions surface. *Virology* **147**:309-316.

Allison, R. F., Sorenson, J. C., Kelly, M. E., Armstrong, F. B., and Dougherty, W. G. 1985b. Sequence determination of the capsid protein gene and flanking regions of tobacco etch virus: evidence for synthesis and processing of a polyprotein in potyvirus genome expression. *Proceedings. National Academy of Sciences USA* **82**:3969-3972.

Alquist, P., French, R., Janda, M., and Loesch-Freis, L. S. 1984. Multicomponent plant virus infection derived from cloned viral cDNA. *Proceedings. National Academy of Sciences USA* **81**:7066-7070.

Alwine, J. C., Kemp, D. J., and Stark, G. R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridisation with DNA probes. *Proceedings. National Academy of Sciences USA* **74**:5350-5354.

Armitage, P., Walden, R., and Draper, J. 1988. Vectors for the transformation of plant cells using *Agrobacterium*. *In Plant Genetic Transformation and Gene Expression: A Laboratory Manual* (Eds., J. Draper, R. Scott, P. Armitage and R. Walden), Pages 1-68. Blackwell Scientific Publications, Oxford.

Atkinson, M. R., Deutscher, M. R., Kornberg, A., Russel, A. F., and Moffat, J. G. 1969. Enzymic synthesis of deoxyribonucleic acid XXXIV. Termination of chain growth by a 2', 3'-dideoxyribonucleotide. *Biochemistry* **8**:4897-4904.

Bankier, A. T., Weston, K. M., and Barrell, B. G. 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods in Enzymology* **155**:51-93.

Barnes, W. M. 1987. Sequencing DNA with dideoxyribonucleotides as chain terminators: hints and strategies for big projects. *Methods in Enzymology* **152**:538-556.

Batteiger, B., Newhall, W. J., and Jones, R. B. 1982. The use of Tween 20 as a blocking agent in the immunological detection of proteins transferred to nitrocellulose membranes. *Journal of Immunological Methods* **55**:297-307.

Baulcombe, D. 1989. Strategies for virus resistance in plants. *Trends in Genetics* **5**:56-60.

Baulcombe, D. C., Saunders, G. R., Bevan, M. W., Mayo, M. A., and Harrison, B. D. 1986. Expression of biologically active viral satellite RNA from the nuclear genome of transformed plants. *Nature* **321**:446-449.

Berger, P. H., and Pirone, T. P. 1986. The effect of helper component on the uptake and localisation of potyviruses in *Myzus persicae*. *Virology* **153**:256-261.

Bevan, M. 1984. Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Research* **12**:8711-8721.

Bevan, M. W., Flavell, R. B., and Chilton, M.-D. 1983. A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature* **304**:184-187.

Bevan, M. W., Mason, S. E., and Goelet, P. 1985. Expression of tobacco mosaic virus coat protein by a cauliflower mosaic virus promoter in plants transformed by *Agrobacterium*. *European Molecular Biology Organisation* **4**:1921-1926.

Birnboim, H. C., and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**:1513-1523.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R., and Klug, A. 1978. Protein disc of tobacco mosaic virus at 2.8 A resolution showing the interactions within and between subunits. *Nature* **276**:362-368.

Bos, L. 1970. The identification of three new viruses isolated from *Wisteria* and *Pisium* in the Netherlands, and the problem of variation within the potato virus Y group. *Netherlands Journal of Plant Pathology* **76**:8-46.

Bradford, M. M. 1976. A rapid and sensitive method for quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* **72**:248-254.

Brakke, M. K., and Van Pelt, N. 1970. Properties of infectious ribonucleic acid from wheat streak mosaic virus. *Virology* **42**:699-706.

Broglie, R., Coruzzi, G., Fraley, R. T., Rogers, S. G., Horsch, R. B., Niedermeyer, J. G., Fink, C. L., Flick, J. S., and Chua, N.-H. 1984. Light-regulated expression of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene in transformed plant cells. *Science* **224**:838-843.

Burton, K., and Petersen, G. B. 1960. The frequencies of certain sequences of nucleotides in deoxyribonucleic acid. *Biochemical Journal* **75**:17-27.

Caplan, A., Herrera-Estrella, L., Inze, D., Van Haute, E., Van Montagu, M., Schell, J., and Zambryski, P. 1983. Introduction of genetic material into plant cells. *Science* **222**:815-821.

Carrington, J. C., Cary, S. M., and Dougherty, W. G. 1988. Mutational analysis of tobacco etch virus polyprotein processing: *cis* and *trans* proteolytic activities of polyproteins containing the 49 kda proteinase. *Journal of Virology* **62**:2313-2320.

Carrington, J. C., Cary, S. M., Parks, T. D., and Dougherty, W. G. 1989. A second proteinase encoded by a plant potyvirus genome. *European Molecular Biology Organisation* **8**:365-370.

Carrington, J. C., and Dougherty, W. G. 1987. Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology* **160**:355-362.

Carrington, J. C., and Dougherty, W. G. 1988. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proceedings. National Academy of Sciences USA* **85**:3391-3395.

Cassells, A. C., and Herrick, C. C. 1977. Cross protection between severe and mild strains of tobacco mosaic virus in doubly inoculated tomato plants. *Virology* **78**:253-260.

Chamberlin, E. E., Atkinson, J. D., and Hunter, J. A. 1964. Cross-protection between strains of apple mosaic virus. *New Zealand Journal of Agricultural Research* **7**:480-490.

Chou, P. Y., and Fasman, G. D. 1974. Prediction of protein conformation. *Biochemistry* **13**:222-245.

Christie, R. G., and Edwardson, J. R. 1977. Light and electron microscopy of plant virus inclusions *(Florida Agricultural Experimental Station Monograph Series*, no. 4). Institute of Food and Agricultural Sciences, University of Florida, Gainsville, USA.

Coleman, A., Byers, M. J., Primrose, S. B., and Lyons, A. 1978. Rapid purification of plasmid DNAs by hydroxyapatite chromatography. *European Journal of Biochemistry* **91**:303-310.

Crouse, J., and Amorese, D. 1987. Ethanol precipitation: ammonium acetate as an alternative to sodium acetate. *BRL Focus* **9**:3-5.

D'Alessio, J. M., Noon, M. C., Ley, H. L., and Gerard, G. 1987. One-tube double -stranded cDNA synthesis using cloned M-MLV reverse transcriptase. *BRL Focus* **9**:1-4.

De Bokx, J. A., and Huttinga, H. 1981. Potato virus Y group. *CMI/AAB Descriptions of Plant Viruses* **no. 242**. (Commonwealth Agricultural Bureaux/Association of Applied Biologists)

Delgado-Sanchez, S., and Grogan, R. G. 1966. Purification and properties of Potato Virus Y. *Phytopathology* **56**:1397-1404.

Dellaporta, S. L., Wood, J., and Hicks, J. B. 1983. A plant DNA minipreparation: version II. *Plant Molecular Biology Reporter* **1**:19-21.

De Mejia, M. V. G., Hiebert, E., and Purcifull, D. E. 1985a. Isolation and partial characterisation of the amorphous cytoplasmic inclusions associated with infections caused by two potyviruses. *Virology* **142**:24-33.

De Mejia, M. V. G., Hiebert, E., Purcifull, D. E., Thornbury, D. W., and Pirone, T. P. 1985b. Identification of potyviral amorphous inclusion protein as a nonstructural, virus-specific protein related to helper component. *Virology* **142**:34-43.

Dente, L., and Cortese, R. 1987. pEMBL: A new family of single stranded plasmids for sequencing DNA. *Methods in Enzymology* **155**:111-119.

De Zoeten, G. A., and Fulton, R. W. 1975. Understanding generates possibilities. *Phytopathology* **65**:221-222.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. 1980. Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proceedings. National Academy of Sciences USA* **77**:7347-7351.

Dodds, J. A. 1982. Cross-protection and interference between electrophoretically distinct strains of cucumber mosaic virus in tomato. *Virology* **118**:235-240.

Dodds, J. A., Lee, S. Q., and Tiffany, M. 1985. Cross-protection between strains of cucumber mosaic virus: effect of host and type of inoculum on accumulation of virions and double stranded RNA of the challenge strain. *Virology* **144**:301-309.

Domier, L. L., Franklin, K. M., Shahabuddin, M., Hellmann, G. M., Overmeyer, J. H., Hiremath, S. T., Slaw, M. F. E., Lomonossoff, G. P., Shaw, J. G., and Rhoads, R. E. 1986. The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Research* **14**:5417-5430.

Domier, L. L., Shaw, J. G., and Rhoads, R. E. 1987. Potyviral proteins share amino acid sequence homology with picorna-, como-, and caulimoviral proteins. *Virology* **158**:20-27.

Dougherty, W. G. 1983. Analysis of viral RNA isolated from tobacco leaf tissue infected with tobacco etch virus. *Virology* **131**:473-481.

Dougherty, W. G., Allison, R. F., Parks, T. D., Johnston, R. E., Feild, M. J., and Armstrong, F. B. 1985. Nucleotide sequence at the 3' terminus of pepper mottle virus genomic RNA: evidence for an alternative mode of potyvirus capsid gene organisation. *Virology* **146**:282-291.

Dougherty, W. G., and Carrington, J. C. 1988. Expression and function of potyviral gene products. *Annual Review of Phytopathology* **26**:123-143.

Dougherty, W. G., and Hiebert, E. 1980a. Translation of potyvirus RNA in a rabbit reticulocyte: reaction conditions and identification of capsid protein as one of the products of *in vitro* translation of tobacco etch and pepper mottle viral RNAs. *Virology* **101**:466-474.

Dougherty, W. G., and Hiebert, E. 1980b. Translation of potyvirus RNA in a rabbit reticulocyte lysate: identification of nuclear inclusion proteins of tobacco etch virus RNA translation and cylindrical inclusion protein as a product of the potyvirus genome. *Virology* **104**:174-182.

Dougherty, W. G., and Hiebert, E. 1980c. Translation of potyvirus RNA in a rabbit reticulocyte lysate: cell-free translation strategy and a genetic map of the potyviral genome. *Virology* **104**:183-194.

Draper, J., Scott, R., and Hamil, J. 1988. Transformation of dicotyledonous plant cells using the Ti plasmid of *Agrobacterium tumefaciens* and the Ri plasmid of *A. rhizogenes*. *In Plant Genetic Transformation and Gene Expression: A Laboratory Manual* (Eds., J. Draper, R. Scott, P. Armitage and R. Walden), Pages 69-160. Blackwell Scientific Publications, Oxford.

Dreher, T. W., and Hall, T. C. 1988. Mutational analysis of the sequence and structural requirements in brome mosaic virus RNA for minus strand promoter activity. *Journal of Molecular Biology* **201**:31-40.

Eck, R. V., and Dayhoff, M. O. 1966. Atlas of protein sequence and structure *(National Biomedical Research Foundation)*, Silver Spring, Maryland, USA.

Edwardson, J. R. 1974. Some properties of the potato virus Y group *(Florida Agricultural Experiment Stations Monograph Series*, no. 4). Institute of Food and Agricultural Sciences, University of Cainsville, USA.

Efstratiadis, A., Kafatos, F., Maxam, A. M., and Maniatis, T. 1976. Enzymic *in vitro* synthesis of globin genes. *Cell* **7**:279-288.

Eggenberger, A. L., Stark, D. M., and Beachy, R. N. 1988. Nucleotide sequence and expression of the soybean mosaic virus coat protein coding region in *Esherichia coli* and tobacco callus. *Virology* **(submitted)**.

Ey, P. L., and Ashman, L. K. 1986. The use of alkaline phosphatase-conjugated anti-immunoglobulin with immunoblots for determining the specificity of monoclonal antibodies to protein mixtures. *Methods in Enzymology* **121**:497-509.

Fellowes, A. 1988. A cDNA clone from a local isolate of PVY^N. M.Sc. Thesis. University of Canterbury, Christchurch, New Zealand.

Felsenstein, J. 1987. PHYLIP (Phylogeny Inference Package) version 3.0 Manual, University of Washington, Washington.

Fitch, W. M. 1971. Towards defining the course of evolution: minimum change for a specified tree topology. *Systematic Zoology* **20**:406-416.

Fletcher, J. 1986. Report to the potato research advisory board. (Crop Research Division, DSIR, Lincoln, New Zealand, rept. no. 117), pp11-13.

Fourney, R. M., Miyakoshi, J., Day, R. S., and Paterson, M. C. 1988. Northern blotting: efficient RNA staining and transfer. *Focus* **10**:5-7.

Fraley, R. T., Rogers, S. G., Horsch, R. B., Sanders, P. R., Flick, J. S., Adams, S. P., Bittner, M. L., Brand, L. A., Fink, C. L., Fry, J. S., Gallupi, G. R., Goldberg, S. B., Hoffmann, N. L., and Woo, S. C. 1983. Expression of bacterial genes in plant cells. *Proceedings. National Academy of Sciences USA* **80**:4803-4807.

Francki, R. I. B. 1983. Current problems in plant virus taxonomy. *In A Critical Appraisal of Viral Taxonomy* (Ed., R. E. F. Matthews), Pages 63-104. CRC Press Inc., Florida, USA.

Francki, R. I. B., Milne, R. G., and Hatta, T. 1985. *Atlas of Plant Viruses, vol.I & II*. CRC Press Inc., Florida, USA.

French, R., and Alquist, P. 1988. Characterisation and engineering of sequences controlling *in vivo* synthesis of brome mosaic virus subgenomic RNA. *Journal of Virology* **62**:2411-2420.

Frenkel, M. J., Ward, C. W., and Shukla, D. D. 1989. Nucleotide sequence of the 3'-untranslated region of the viral genome as a basis for the identification and classification of potyviruses: its application to watermelon mosaic virus 2 and soybean mosaic virus-N. *Journal of General Virology* **(submitted)**.

Fulton, R. W. 1980. The protective effects of systemic virus infection. *In NATO Advanced Study Institute on Active Defense Mechanisms in Plants* (Ed., R. K. S. Wood). Volume 37, Pages 231-245. Plenum Press, New York, USA.

Fulton, R. W. 1986. Practices and precautions in the use of cross protection for plant virus disease control. *Annual Review of Plant Pathology* **24**:67-81.

Gallie, D. R., Lucas, W. J., and Walbot, V. 1989. Visualising mRNA expression in plant protoplasts: factors influencing efficient mRNA uptake and translation. *The Plant Cell* **1**:301-311.

Garfinkel, D. J., and Nester, E. W. 1980. *Agrobacterium tumefaciens* mutants affected in crown gall tumorigenesis and octopine catabolism. *Journal of Bacteriology* **144**:732-743.

Garnier, J. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* **120**:97-120.

Gibbs, A. J. 1969. Plant virus classification. *Advances in Virus Research* **14**:263-328.

Gibbs, A. J. 1977. Tobamovirus group. *CMI/AAB Descriptions of Plant Viruses* **no.184**. (Commonwealth Agricultural Bureaux/Association of Applied Biologists)

Gough, K. H., Azad, A. A., Hanna, P. J., and Shukla, D. D. 1987. Nucleotide sequence of the capsid and nuclear inclusion protein genes from the Johnson grass strain of sugarcane mosaic virus RNA. *Journal of General Virology* **68**:297-304.

Gough, K. H., and Shukla, D. D. 1981. Coat protein of potyviruses. 1. Comparison of the four Australian strains of sugarcane mosaic virus. *Virology* **111**:455-462.

Gronenborn, B., and Messing, J. 1978. Methylation of single-stranded DNA *in vitro* introduces new restriction endonuclease cleavage sites. *Nature* **272**:375-377.

Gubler, U., and Hoffman, B. J. 1983. A simple and very efficient method for generating cDNA libraries. *Gene* **25**:263-269.

Gunyuzlu, P. L., Tolin, S. A., and Johnson, J. L. 1987. The nucleotide sequence of the 3'-terminus of soyabean mosaic virus. *Phytopathology* **77**:1766.

Habili, N., McInnes, J. L., and Symons, R. H. 1987. Nonradioactive photobiotin labelled DNA probes for the routine diagnosis of barley yellow dwarf virus. *Journal of Virological Methods* **16**:225-237.

Hamilton, R. I. 1980. Defences triggered by previous invaders-viruses. *In Plant Disease* (Eds., J. G. Horsfall and E. B. Cowling). Volume 5, Pages 279-303. Academic Press, Inc., New York, USA.

Hamilton, R. I., and Nichols, C. 1978. Serological methods for detection of pea seed borne mosaic virus in leaves and seeds of *Pisium sativum*. *Phytopathology* **68**:539-543.

Hamlyn, P. H., Brownlee, G. G., Chen, C., Gait, M. J., and Milstein, C. 1978. Complete sequence of constant and 3' non-coding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing. *Cell* **15**:1067-1075.

Hanahan, D. 1983. Studies in the transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**:557-580.

Hari, V. 1981. The RNA of tobacco etch virus: further characterisation and detection of protein linked to RNA. *Virology* **112**:391-399.

Hari, V., Siegel, A., Rozek, C., and Timberlake, W. E. 1979. The RNA of tobacco etch virus contains poly(A). *Virology* **92**:568-571.

Harpster, M. H., Townsend, J. A., Jones, J. D. G., Bedbrook, J., and Dunsmuir, P. 1988. Relative strengths of the 35S cauliflower mosaic virus, 1', 2', and nopaline synthase promoters in transformed tobacco, sugarbeet and oilseed rape callus tissue. *Molecular and General Genetics* **212**:182-190.

Harrison, B. D. 1985. Usefulness and limitations of the species concept for plant viruses. *Intervirology* **25**:71-78.

Harrison, B. D., Finch, J. T., Gibbs, A. J., Hollings, M., Shepherd, R. J., Valenta, V., and Wetter, C. 1971. Sixteen groups of plant viruses. *Virology* **45**:356-363.

Hay, J. M., Fellowes, A. P., and Timmerman, G. M. 1989. Nucleotide sequence for the coat protein gene of a New Zealand isolate of potato virus Y, necrotic strain. *Archives of Virology* **(in press)**.

Heidecker, G., Messing, J., and Gronenborn, B. 1980. A versatile primer for DNA sequencing in the M13mp2 cloning system. *Gene* **10**:69-73.

Hellmann, G. M., Hiremath, S. T., Shaw, J. G., and Rhoads, R. E. 1986. Cistron mapping of tobacco vein mottling virus. *Virology* **151**:159-171.

Hellmann, G. M., Shaw, J. G., Lesnaw, J. A., Chu, L.-Y., Pirone, T. P., and Rhoads, R. E. 1980. Cell-free translation of tobacco vein mottling virus RNA. *Virology* **106**:207-216.

Hellmann, G. M., Shaw, J. G., and Rhoads, R. E. 1985. On the origin of the helper component of tobacco vein mottling virus: translation initiation near the 5'-terminus of the viral RNA and termination by UAG codons. *Virology* **143**:23-34.

Hellmann, G. M., Shaw, J. G., and Rhoads, R. E. 1987. *In vitro* analysis of tobacco vein mottling virus Nla cistron: evidence for a virus-encoded protease. *Virology* **163**:554-562.

Hellmann, G. M., Thornbury, D. W., Hiebert, E., Shaw, J. G., Pirone, T. P., and Rhoads, R. E. 1983. Cell-free translation of tobacco vein mottling virus RNA. II Immunoprecipitation of products by antisera to cylindrical inclusion, nuclear inclusion and helper component proteins. *Virology* **124**:434-444.

Hemenway, C., Fang, R.-X., Kaniewski, W. K., Chua, N.-H., and Tumer, N. E. 1988. Analysis of the mechanism of protection in transgenic plants expressing the potato virus X coat protein or its antisense RNA. *European Molecular Biology Organisation* **7**:1273-1280.

Hepburn, A. G., Clarke, L. E., Pearson, L., and White, J. 1983. The role of cytosine methylation in the control of nopaline synthase expression in a plant tumor. *Journal of Molecular and Applied Genetics* **2**:315-329.

Herrera-Estrella, L., De Block, M., Messens, E., Hernalsteens, J.-P., Van Montagu, M., and Schell, J. 1983. Chimaeric genes as selectable markers in plant cells. *European Molecular Biology Organisation* **2**:987-995.

Herrera-Estrella, L., and Simpson, J. 1988. Foreign gene expression in plants. *In Plant Molecular Biology: A Practical Approach* (Ed., C. H. Shaw), Pages 131-160. IRL Press, Oxford, UK.

Herrera-Estrella, L., Van den Broeck, G., Maenhaut, R., Van Montagu, M., Schell, J., Timko, M., and Cashmore, A. 1984. Light-inducible and chloroplast-associated expression of a chimaeric gene introduced into *Nicotiana tabacum* using a Ti plasmid. *Nature* **310**:115-120.

Herrmann, R., Neugebauger, K., Pirkl, E., Zentgraf, H., and Schaller, H. 1980. Single-stranded DNA phage fd as a cloning vector. *Molecular and General Genetics* **177**:231-242.

Hiebert, E., and McDonald, J. G. 1973. Characterisation of some proteins associated with viruses in the potato Y group. *Virology* **56**:349-361.

Hiebert, E., Thornbury, D. W., and Pirone, T. P. 1984. Immunoprecipitation analysis of potyviral *in vitro* translation products using antisera to helper component of tobacco vein mottling virus and potato virus Y. *Virology* **135**:1-9.

Hill, J. H., and Benner, H. I. 1976. Properties of potyvirus RNAs: turnip mosaic, tobacco etch and maize dwarf mosaic viruses. *Virology* **75**:419-432.

Hill, J. H., Ford, R. E., and Benner, H. I. 1973. Purification and partial characterisation of maize dwarf mosaic virus strain B (sugarcane mosaic virus). *Journal of General Virology* **20**:327-339.

Hill, J. H., and Shepherd, R. J. 1972. Biochemical properties of turnip mosaic virus. *Virology* **47**:807-816.

Hindley, J. 1983. *DNA sequencing. Laboratory techniques in biochemistry and molecular biology*. Elsevier Biomedical Press, Oxford. (Work,TS; Burdon,RH: Eds)

Hinostroza-Orihuela, A. M. 1975. Some characteristics of infectious RNA from Potato Virus Y. *Virology* **67**:276-278.

Hoekema, A., Hirsh, P. R., Hooykaas, P. J. J., and Schilperoot, R. A. 1983. A binary plant vector strategy based on separation of the *vir*- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* **303**:179-180.

Hollings, M., and Brunt, A. A. 1981. Potyviruses. *In Handbook of Plant Virus Infections and Comparative Diagnosis* (Ed., E. Kurstak), Pages 731-807. Elesvier/North Holland Biomedical Press.

Hopp, T. P., and Woods, K. R. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proceedings. National Academy of Sciences USA* **78**:3824-3828.

Horsch, R. B., Fraley, R. T., Rogers, S. G., Sanders, P. R., Lloyd, A., and Hoffmann, N. 1984. Inheritance of functional foreign genes in plants. *Science* **223**:496-498.

Horsch, R. B., Fry, J. E., Hoffmann, N. L., Eichholtz, D., Rogers, S. G., and Fraley, R. T. 1985. A simple and general method for transferring genes into plants. *Science* **227**:1229-1231.

Hull, R., and Davies, J. W. 1983. Genetic engineering with plant viruses, and their potential as vectors. *Advances In Virus Research* **28**:1-33.

Jordon, M. C., and McHughen, A. 1988. Transformed callus does not necessarily regenerate transformed shoots. *Plant Cell Reports* **7**:285-287.

Klee, H. J., Yanofsky, M. F., and Nester, E. W. 1985. Vectors for transformation of higher plants. *Biotechnology* **3**:637-642.

Klinowski, M., and Schmelzer, K. 1960. A necrotic type of potato virus Y. *American Potato Journal* **37**:221-228.

Knesek, J. E., Mink, G. I., and Hampton, R. O. 1974. Purification and properties of pea seed-borne mosaic virus. *Phytopathology* **64**:1076-1081.

Knutsen, H., Hiebert, E., and Purcifull, D. E. 1974. Partial purification and some properties of tobacco etch virus induced intranuclear inclusions. *Virology* **61**:200-209.

Koenig, R., and Lesemann, D. E. 1979. Tymovirus group. *CMI/AAB Descriptions of Plant Viruses* **no. 214**. (Commonwealth Agricultural Bureaux/Association of Applied Biologists)

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**:283-292.

Kyte, J., and Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**:105-132.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* **227**:680-685.

Loesch-Fries, L. S., Merlo, D., Zinnen, T., Hill, L., Krahn, K., Jarvis, N., Nelson, S., and Halk, E. 1987. Expression of alfalfa mosaic virus RNA4 in transgenic plants confers virus resistance. *European Molecular Biology Organisation* **6**:1845-1851.

Makkouk, K. M., and Gumpf, D. J. 1975. Characterisation of the protein and nucleic acid of potato virus Y strains isolated from pepper. *Virology* **63**:336-344.

Maliga, P. 1984. Cell culture procedures for mutant selection and characterisation in *Nicotiana plumbaginifolia*. *In Cell Culture and Somatic Cell Genetics of Plants* (Ed., I. K. Basil), *Laboratory procedures and their implications*. Volume 1, Pages 552-562. Academic Press, Orlando.

Mandel, M., and Higa, A. 1970. Calcium-dependent bacteriophage DNA infection. *Journal of Molecular Biology* **53**:159-165.

Maniatis, T., Fritsch, E. F., and Sambrook, J. 1982. *Molecular cloning: a laboratory manual. Cold Spring Harbour laboratory*. Cold Spring Harbour, New York.

Matthews, R. E. F. 1981. The groups of plant viruses. *In Plant Virology*, 2nd EditionPages 715-725. Academic Press, London.

Maxam, A., and Gilbert, W. 1977. A new method for sequencing DNA. *Proceedings. National Academy of Sciences USA* **74**:560-564.

McDonald, J. G., and Bancroft, J. B. 1977. Assembly studies on potato virus Y and its coat protein. *Journal of General Virology* **35**:251-263.

McDonald, J. G., Beveridge, T. J., and Bancroft, J. B. 1976. Self-assembly of protein from a flexuous virus. *Virology* **69**:327-333.

McGadey, J. 1970. A tetrazolium method for non-specific alkaline phosphatase. *Histochemie* **23**:180-184.

McKinney, H. H. 1929. Mosaic disease in the Canary Islands, West Africa and Gilbraltar. *Journal of Agricultural Research* **39**:557-578.

Meshi, T., Ishikawa, M., Motoyoshi, F., Semba, K., and Okada, Y. 1986. *In vitro* transcription of infectious RNAs from full-length cDNAs of tobacco mosaic virus. *Proceedings. National Academy of Sciences USA* **83**:5043-5047.

Messing, J., Crea, R., and Seeburg, P. H. 1981. A system for shot-gun DNA sequencing. *Nucleic Acids Research* **9**:309-321.

Messing, J., Gronenborn, B., Muller-Hill, B., and Hofschneider, P. H. 1977. Filamentous coliophage M13 as a cloning vehicle: insertion of a *Hind*III fragment of the *lac* regulatory region in M13 replicative form *in vitro*. *Proceedings. National Academy of Sciences USA* **74**:3642-3646.

Messing, J., and Viera, J. 1982. A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. *Gene* **19**:269-276.

Mierendorf, R. C., and Pfeffer, D. 1987. Direct sequencing of denatured plasmid DNA. *Methods in Enzymology* **152**:556-566.

Moghul, S. M., and Francki, R. I. B. 1976. Towards a system for the identification and classification of potyviruses. 1. Serology and amino acid composition of six distinct viruses. *Virology* **73**:350-362.

Moore, L., Warren, G., and Strobel, G. 1979. Involvement of a plasmid in the hairy root disease of plants caused by *Agrobacterium rhizogenes*. *Plasmid* **2**:617-626.

Mount, D. W., and Conrad, B. 1986. Improved programs for DNA and protein sequence analysis on the IBM personal computer and other computer systems. *Nucleic Acids Research* **14**:442-445.

Murashige, T., and Skoog, F. 1962. A revised method for rapid growth and bioassays with tissue cultures. *Physiologia Plantarum* **15**:473-497.

Murray, E. E., Lotzer, J., and Eberle, M. 1989. Codon usage in plant genes. *Nucleic Acids Research* **17**:477-498.

Nelson, M. R., Wheeler, R. E., and Zitter, T. A. 1982. Pepper Mottle Virus. *CMI/AAB Descriptions of Plant Viruses* **no.253**. (Commonwealth Agricultural Bureaux/Association of Applied Biologists)

Nelson, R. S., Powell Abel, P., and Beachy, R. 1987. Lesions and virus accumulation in inoculated transgenic tobacco plants expressing the coat protein gene of Tobacco Virus Virus. *Virology* **158**:126-132.

Oakley, B. R., Kirsch, D. R., and Morris, N. R. 1980. A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Analytical Biochemistry* **105**:361-363.

Okayama, H., and Berg, P. 1982. High-efficiency cloning of full-length cDNA. *Molecular and Cellular Biology* **2**:161-170.

Ooms, G., Karp, A., Burrell, M. M., Twell, D., and Roberts, J. 1985. Genetic modification of potato development using Ri T-DNA. *Theoretical and Applied Genetics* **70**:440-446.

Ooms, G., Karp, A., and Roberts, J. 1983. From tumour to tuber; tumour cell characteristics and chromosome numbers of crown gall-derived tetraploid potato plants (*Solanum tuberosum* cv. 'Maris Bard'). *Theoretical and Applied Genetics* **66**:169-172.

Palukaitis, P., and Zaitlin, M. 1984. A model to explain the 'cross protection' phenomenon shown by plant viruses and viroids. *In Plant-Microbe Interactions: Molecular and Genetic Perspectives* (Eds., T. Kosuge and E. W. Nester), Volume 1, Pages 421-429. MacMillan, New York.

Pfeffer, D., and Mierendorf, R. 1986. Direct sequencing and pGEM plasmids: klenow versus reverse transcriptase. *Promega notes* **5**. (Promega).

Pirone, T. P., and Thornbury, D. W. 1983. Role of virion and helper component in regulating aphid transmission of tobacco etch virus. *Phytopathology* **73**:872-874.

Pohl, F., and Beck, S. 1987. Direct transfer electrophoresis used for DNA sequencing. *Methods in Enzymology* **155**:250-260.

Powell Abel, P., Nelson, R. S., De, B., Hoffmann, N., Rogers, S. G., Fraley, R. T., and Beachy, R. N. 1986. Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. *Science* **232**:738-743.

Rancaniello, V. R., and Baltimore, D. 1981. Cloned poliovirus complementary DNA is infectious in mammalian cells. *Science* **214**:916-918.

Rast, A. T. B. 1972. MII-16, an artificial symptomless mutant of tobacco mosaic for seedling inoculation of tomato crops. *Netherlands Journal of Plant Pathology* **78**:110-112.

Ravelonandro, M., Varveri, C., Delbos, R., and Dunez, J. 1988. Nucleotide sequence of the capsid protein gene of plum pox potyvirus. *Journal of General Virology* **69**:1509-1516.

Reddick, B. B., and Barnett, O. W. 1983. A comparison of three potyviruses by direct hybridisation analysis. *Phytopathology* **73**:1506-1510.

Rosenthal, A., Jung, R., and Hunger, H.-D. 1987. Solid-phase methods for sequencing of oligodeoxy ribonucleotides and DNA. *Methods in Enzymology* **155**:301-331.

Rosner, A., and Bar-Joseph, M. 1984. Diversity of citrus tristeza virus strains indicated by hybridisation with cloned cDNA sequences. *Virology* **139**:189-193.

Rosner, A., Raccah, B., Mayoral, M. L., Bar-Joseph, M., and Ginzburg, I. 1986. Synthesis of DNA complementary to the polyadenylated genomic RNA of potato virus Y and its molecular cloning. *Plant Pathology* **35**:178-184.

Ross, A. F. 1974. Interactions of viruses in the host. *In Virus diseases of ornamental hosts* (Eds., R. H. Lawson and M. K. Corbett), Pages 247-260. Society of horticultural science, The Hague.

Rougeon, F., and Mach, B. 1976. Step-wise biosynthesis *in vitro* of globin genes from globin mRNA by DNA polymerase of avian myeloblastosis virus. *Proceedings. National Academy of Sciences USA* **73**:3418-3422.

Salaman, R. N. 1933. Protective inoculation against a plant virus. *Nature* **131**:468.

Sanger, F., and Coulson, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **94**:441-448.

Sanger, F., and Coulson, A. R. 1978. The use of thin acrylamide gels for DNA sequencing. *FEBS letters (Federation of European Biochemical Societies)* **87**:107-110.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**:161-178.

Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H., and Fischer, D. 1973. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage F1 DNA. *Proceedings. National Academy of Sciences USA* **70**:1209-1213.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings. National Academy of Sciences USA* **74**:5463-5467.

Sawyer, L., Tollin, P., and Wilson, H. R. 1987. A comparison between the predicted secondary structures of potato virus X and papaya mosaic virus coat proteins. *Journal of General Virology* **68**:1229-1232.

Scott, R. 1988. DNA restriction and analysis by Southern hybridisation. *In Plant Genetic Transformation and Gene Expression: A Laboratory Manual* (Eds., J. Draper, R. Scott, P. Armitage and R. Walden), Pages 237-265. Blackwell Scientific Publications, Oxford.

Scott, R., Draper, J., Jefferson, R., Dury, G., and Jacob, L. 1988. Analysis of gene organisation and expression in plants. *In Plant Genetic Transformation and Gene Expression: A Laboratory Manual* (Eds., J. Draper, R. Scott, P. Armitage and R. Walden), Pages 263-339. Blackwell Scientific Publications, Oxford.

Sequeira, L. 1984. Cross protection and induced resistance: their potential for plant disease resistance. *Trends in Biotechnology* **2**:25-29.

Shah, D. M., Horsch, R. B., Klee, H. J., Kishorer, G. M., Winter, J. A., Tumer, N. E., Hironaka, C. M., Sanders, P. R., Gasser, C. S., Aykent, S., Siegel, N. R., Rogers, S. G., and Fraley, R. T. 1986. Engineering herbicide tolerance in transgenic plants. *Science* **233**:478-481.

Shahabuddin, M., Shaw, J. G., and Rhoads, R. E. 1988. Mapping of the tobacco vein mottling virus VPg cistron. *Virology* **163**:635-637.

Shahin, E. A., and Simpson, R. B. 1986. Gene transfer and system for potato. *HortScience* **21**:1199-1201.

Shalla, T. A., and Peterson, L. J. 1978. Studies on the mechanism of viral cross protection. *Phytopathology* **68**:1681-1683.

Sherwood, J. L., and Fulton, R. W. 1982. The specific involvement of coat protein in tobacco mosaic virus cross protection. *Virology* **119**:150-158.

Sherwood, J. L., and Fulton, R. W. 1983. Competition for infection sites and multiplication of the competing strain in plant viral interference. *Phytopathology* **73**:1363-1365.

Shukla, D. D., Gough, K. H., and Ward, C. W. 1987. Coat protein of potyviruses. 3. Comparison of amino acid sequences of the coat proteins of four Australian strains of sugarcane mosaic virus. *Archives of Virology* **96**:59-74.

Shukla, D. D., Inglis, A. S., McKern, N. M., and Gough, K. H. 1986. Coat protein of potyviruses. 2. Amino acid sequence of the coat protein of potato virus Y. *Virology* **152**:118-125.

Shukla, D. D., Jilka, J., Tosic, M., and Ford, R. E. 1989. A novel approach to the serology of potyviruses involving affinity-purified polyclonal antibodies directed towards virus-specific amino termini of coat proteins. *Journal of General Virology* **70**:13-23.

Shukla, D. D., McKern, N. M., Gough, K. H., Tracey, S. L., and Letho, S. G. 1988a. Differentiation of potyviruses and their strains by high performance liquid chromatographic peptide profiling of coat proteins. *Journal of General Virology* **69**:493-502.

Shukla, D. D., McKern, N. M., and Ward, C. W. 1988b. Coat protein of potyviruses. 5. Symptomatology, serology and coat protein sequences of three strains of passionfruit woodiness virus. *Archives of Virology* **102**:221-232.

Shukla, D. D., Strike, P. M., Tracey, S. L., Gough, K. H., and Ward, C. W. 1988c. The N and C termini of the coat proteins of potyviruses are surface located and the N terminus contains the major virus-specific epitopes. *Journal of General Virology* **69**:1497-1508.

Shukla, D. D., Thomas, J. E., McKern, N. M., Tracey, S. L., and Ward, C. W. 1988d. Coat protein of potyviruses. 4. Comparison of biological properties, serological relationships and coat protein amino acid sequences of four strains of potato virus Y. *Archives of Virology* **102**:207-220.

Shukla, D. D., Tosic, M., Jilka, J., Ford, R. E., Toler, R. W., and Langham, M. A. C. 1988e. Taxonomy of potyviruses infecting maize, sorgham and sugarcane in Australia and the United States as determined by reactivities of polyclonal antibodies directed towards virus-specific amino-termini of coat proteins. *Phytopathology* **(in press)**.

Shukla, D. D., and Ward, C. W. 1988a. Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. *Advances in Virus Research* **(in press)**.

Shukla, D. D., and Ward, C. W. 1988b. Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *Journal of General Virology* **69**:2703-2710.

Shukla.DD, and Ward, C. W. 1989. Identification and classification of potyviruses on the basis of coat protein sequence data and serology. *Archives of Virology* **(in press)**.

Sievert, R. C. 1978. Effect of PVY and TMV on field-grown Burley tobacco. *Phytopathology* **68**:823-825.

Smith, A. J. H. 1979. The use of Exonuclease III for preparing single stranded DNA for use as a template in the chain termination sequencing methods. *Nucleic Acids Research* **6**:831-848.

Smith, K. M., and Dennis, R. W. G. 1940. Some notes on a suspected variant of Solanum virus 2 (potato virus Y). *Annals of Applied Biology* **27**:65-70.

Smith, L., Kaiser, R., Sanders, J., and Hood, L. 1987. The synthesis and use of fluorescent oligonucleotides in DNA sequence analysis. *Methods in Enzymology* **155**:260-301.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**:503-517.

Stace-Smith, R., and Tremaine, J. H. 1970. Purification and composition of potato virus Y. *Phytopathology* **60**:1785-1789.

Staden, R. 1986. The current status and portability of our sequence handling software. *Nucleic Acids Research* **14**:217-231.

Switzer, R. C., Merril, C. R., and Shifrin, S. 1979. A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Analytical Biochemistry* **98**:231-237.

Taniguchi, T., Palmieri, M., and Weissmann, C. 1978. Qβ DNA-containing hybrid plasmids giving rise to Qβ phage formation in the bacterial host. *Nature* **274**:223-228.

Thomashow, M. F., Nutter, R., Postle, K., Chilton, M.-D., Blattner, F. R., Powell Abel, A., Gordon, M. P., and Nester, E. W. 1980. Recombination between higher plant DNA and the Ti plasmid of *Agrobacterium tumefaciens*. *Proceedings. National Academy of Sciences USA* **77**:6448-6452.

Thornbury, D. W., Hellmann, G. M., Rhoads, R. E., and Pirone, T. P. 1985. Purification and characterisation of potyvirus helper component. *Virology* **144**:260-267.

Thornbury, D. W., and Pirone, T. P. 1983. Helper components of two potyviruses are serologically distinct. *Virology* **125**:487-490.

Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. 1973. Improved estimation of secondary structure in ribonucleic acids. *Nature: New Biology* **246**:40-41.

Towbin, H., Staehelin, T., and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings. National Academy of Sciences USA* **76**:4350-4354.

Tumer, N. E., O'Connell, K. M., Nelson, R. S., Sanders, P. R., Beachy, R. N., Fraley, R. T., and Shab, D. M. 1987. Expression of alfalfa mosaic virus coat protein gene confers cross protection in transgenic tobacco and tomato plants. *European Molecular Biology Organisation* **6**:1181-1188.

Turpen, T. 1989. Molecular cloning of a potato virus Y genome. **(submitted)**.

Valle, R. P. C., Skrzeczkowski, J., Morch, M.-D., Joshi, R. L., Gargouri, R., Drugeon, G., Boyer, J. C., Chapeville, F., and Haenni, A.-L. 1988. Plant viruses and new perspectives in cross protection. *Biochimie* **70**:695-703.

Vance, V. B., and Beachy, R. N. 1984. Translation of soybean mosaic virus RNA *in vitro*: evidence of protein processing. *Virology* **132**:271-281.

Van der Vlugt, R., Allefs, S., De Haan, P., and Goldbach, R. 1989. Nucleotide sequence of the 3'-terminal region of potato virus Y^N RNA. *Journal of General Virology* **70**:229-233.

Van Dun, C. M. P., and Bol, J. F. 1988. Transgenic tobacco plants accumulating tobacco rattle virus coat protein resist infection with tobacco rattle virus and pea early browning virus. *Virology* **167**:649-652.

Van Dun, C. M. P., Bol, J. F., and Van Vloten-Doting, L. 1987. Expression of alfalfa mosaic virus and tobacco rattle virus coat protein genes in transgenic plants. *Virology* **159**:299-305.

Van Dun, C. M. P., Overduin, B., Van Vloten-Doting, L., and Bol, J. F. 1988a. Transgenic tobacco expressing tobacco streak virus or mutated alfalfa mosaic virus coat protein does not cross protect against alfalfa mosaic virus infection. *Virology* **164**:383-389.

Van Dun, C. M. P., Van Vloten-Doting, L., and Bol, J. F. 1988b. Expression of alfalfa mosaic virus cDNA1 and 2 in transgenic tobacco plants. *Virology* **163**:572-578.

Van Haute, E., Joos, H., Maes, S., Warren, G., Van Montagu, M., and Schell, J. 1983. Intergenic transfer and exchange recombination of restriction fragments cloned in pBR322: a novel strategy for reversed genetics of the Ti plasmids of *Agrobacterium tumefaciens*. *European Molecular Biology Organisation* **2**:411-418.

Vieira, J., and Messing, J. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**:259-268.

Vos, P., Verver, J., Van Wezenbeek, P., Van Kammen, A., and Goldbach, R. 1984. Study of the genetic organisation of a plant viral RNA genome by *in vitro* expression of a full-length DNA copy. *European Molecular Biology Organisation* **3**:3049-3053.

Wallace, R., Johnson, M. J., Suggs, S. V., Miyoshi, K., Bhatt, R., and Hakura, K. 1981. A set of synthetic oligodeoxynucleotide primers for DNA sequencing in the plasmid vector pBR322. *Gene* **16**:21-26.

Wang, K., Herrera-Estrella, L., Van Montagu, M., and Zambryski, P. 1984. Right 25 base pair terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from *Agrobacterium* to the plant genome. *Cell* **38**:455-462.

Ward, C. W., and Dopheide, T. A. 1980. Influenza virus haemagglutinin. Structural predictions suggest that the fibrillar appearance is due to the presence of a coiled-coil. *Australian Journal of Biological Sciences* **33**:441-447.

Williams, M. K., Conner, A. J., Gardner, R. C., Deroles, S. C., and Lammerink, J. P. 1988. Gene transfer to potatoes using *Agrobacterium*. *In Ninth Australian Plant Breeding Conference, Proceedings* (Eds., K. S. McWhirter, R. W. Downes and B. J. Rhead), Pages 71-72. Agricultural Research Institute, Waggawagga.

Wu, R. 1987. Preface. *Methods in Enzymology* **155**:xiii.

Yamaya, J., Yoshioka, M., Meshi, T., Okada, Y., and Ohno, T. 1988. Expression of tobacco mosaic virus RNA in transgenic plants. *Molecular and General Genetics* **211**:520-525.

Yarwood, C. E. 1968. Sequence of supplements in virus inoculations. *Phytopathology* **58**:132-136.

Yeh, S. D., and Gonsalves, D. 1985. Translation of papaya ringspot virus RNA *in vitro*: detection of a possible polyprotein that is processed for capsid protein, cylindrical-inclusion protein, and amorphous inclusion protein. *Virology* **143**:260-271.

Yeh, S.-D., Gonsalves, D., Wang, H.-L., Namba, R., and Chiu, R.-J. 1988. Control of papaya ringspot virus by cross protection. *Plant Disease* **72**:375-380.

Zaitlin, M. 1976. Viral cross protection: more understanding is needed. *Phytopathology* **66**:382-383.

Zinnen, T. M., and Fulton, R. W. 1986. Cross protection between sunn-hemp mosaic and tobacco mosaic viruses. *Journal of General Virology* **67**:1679-1687.

APPENDIX A: TERMINATION MIXES FOR DNA SEQUENCING

Following are the final concentrations of the components in each of the four termination mixes.

d/ddT mix: 3.3 μM dTTP, 36.3 μM dGTP, 36.3 μM dCTP, 3.3 μM-83 μM ddTTP, 10 μCi [α-^{35}S]dATP.

d/ddG mix: 1.7 μM dGTP, 36.3 μM dTTP, 36.3 μM dCTP, 11.0 μM-33.3 μM ddGTP, 10 μCi [α-^{35}S]dATP.

d/ddC mix: 1.7 μM dCTP, 36.3 μM dTTP, 36.3 μM dGTP, 6.7 μM-16.7 μM ddCTP, 10 μCi [α-^{35}S]dATP.

d/ddA mix: 25.6 μM dTTP, 25.6 μM dGTP, 25.6 μM dCTP, 4.2 μM-11 μM ddATP, 10 μCi [α-^{35}S]dATP.

APPENDIX B: PVY^N COAT PROTEIN GENE CONSENSUS SEQUENCE

The final compilation DNA consensus **(CON)** sequence for the PVY^N gene is presented. The sequences of the individual fragments generated by endonuclease digestion of the whole gene are included. The restriction endonucleases are abbreviated as follows: *Xba*I (XB), *Alu*I (AL), *Taq*I (TQ), *Rsa*I (RS), *Hpa*II (HP) and *Hae*III (HA). A 'C' in the fragment name indicates the use of the complementary DNA sequence in the consensus sequence. The numbers 1, 2, 3 and 4 indicate bases which have been tentatively assigned as C, T, A and G, respectively. An 'X' denotes a base of unknown identity and a 'V' denotes a possible 'GG'.

```
                10         20         30         40         50         60
(XB.A8)   CACAATCGAT GCAGGAGGAA GCACTAAAAA GGATGCAAAA CAAGAGCAAG GTAGCATTCA
(TQ.24C)       TCGAT GCAGGAGGAA GCACTAAAAA GGATGCAAAA CAAGAGCAAG GTAGCATTCA
(AL.13A)  CACAATCGAT GCAGGAGGAA GCACTAAAAA GGATGCAAAA CAAGAGCAAG GTAGCATTCA
(CON)     CACAATCGAT GCAGGAGGAA GCACTAAAAA GGATGCAAAA CAAGAGCAAG GTAGCATTCA

                70         80         90        100        110        120
(TQ.24C)  ACCAAATTTC AACAAGGAAA AGGAAAAGGA CGTGAATGTT GGAACATCTG GAACTCATAC
(XB.A8)   ACCAAATTTC AACAAGGAAA AGGAAAAGGA CGTGAATGTT GGAACATCTG GAACTCATAC
(AL.13A)  ACCAAATTTC AACAAGGAAA AGGAAAAGGA CGTGAATGTT GGAACATCTG GAACTCATAC
(CON)     ACCAAATTTC AACAAGGAAA AGGAAAAGGA CGTGAATGTT GGAACATCTG GAACTCATAC

              130        140        150        160        170        180
(TQ.24C)  TGTGCCACGA ATTAAAGCTA TCACGTCCAA AATGAGAATG CCCAAGAGTA AAGGTGCAAT
(AL.13A)  TGTGCCACGA ATTAAAG
(AL.X14)                   CTA TCACGTCCAA AATGAGAATG CCCAAGAGTA AAGGTGCAAT
(XB.A8)   TGTGCCACGA ATTAAAGCTA TCACGTCCAA AATGAGAATG CCCAAGAGTA AAGGTGCAAT
(AL.X2)                    CTA TCACGTCCAA AATGAGAATG CCCAAGAGTA AAGGTGCAAT
(CON)     TGTGCCACGA ATTAAAGCTA TCACGTCCAA AATGAGAATG CCCAAGAGTA AAGGTGCAAT

              190        200        210        220        230        240
(AL.23C)                                        ACAG CAAATTGACA TCTCAAATAC
(TQ.X4)                               GAGTA TGCTCCACAG CAAATTGACA TCTCAAATAC
(TQ.X14)                              GAGTA TGCTCCACAG CAAATTGACA TCTCAAATAC
(AL.X14)  TGCATTAAAT TTGGAACACT TACTCGAGTA TGCTCCACAG CAAATTGACA TCTCAAATAC
(TQ.24C)  TGC
(XB.A8)   TGCATTAAAT TTGGAACACT TACTCGAGTA TGCTCCACAG XAAATTGACA TXTXXAXT
(AL.X2)   TGCATTAAAT TTGGAACACT TACTCGAGTA TGCTCCACAG CAAATTGACA TCTCAAATAC
(CON)     TGCATTAAAT TTGGAACACT TACTCGAGTA TGCTCCACAG CAAATTGACA TCTCAAATAC

              250        260        270        280        290        300
(AL.X14)  TCGAGCAACT CAATCACAGT TTGATACGTG GTATGAAGCA GTACAACTTG CATACGACAT
(AL.X2)   TCGAGCAACT CAATCACAGT TTGATACGTG GTATGAAGCA GTACAACTTG CATA1GACAT
(RS.A2)                                                ACAACTTG CATACGACAT
(AL.23C)  TCGAGCAACT CAATCACAGT TTGATACGTG GTATGAAGCA GTACAACTTG CATACGACAT
(TQ.X14)  TCGA
(TQ.X4)   TCGA
(CON)     TCGAGCAACT CAATCACAGT TTGATACGTG GTATGAAGCA GTACAACTTG CATACGACAT
```

```
                310         320         330         340         350         360
(AL.X14)  AGGAGAAACT GAAATGCCAA CTGTGATGAA TGGGCTTATG GTTTGGTGCA TTGAAAATGG
(AL.X2)   AGGAGAAACT GAAATGCCAA CTGTGATGAA TGGGCTTATG GTTTGGTGCA TTGAAAATGG
(RS.A2)   AGGAGAAACT GAAATGCCAA CTGTGATGAA TGGGCTTATG GTTTGGTGCA TTGAAAATGG
(AL.23C)  AGGAGAAACT GAAATGCCAA CTGTGATGAA TGGGCTTATG GTTTGGTGCA TTGAAAATGG
(CON)     AGGAGAAACT GAAATGCCAA CTGTGATGAA TGGGCTTATG GTTTGGTGCA TTGAAAATGG

                370         380         390         400         410         420
(AL.X14)  AACCTCGCCA AACATCAACG GAGTTTGGGT TATGATGGAT GGAGATGAAC AAGTCGAATA
(AL.X2)   AACCTCGCCA AACATCAACG GAGTTTGGGT TATGATGGAT GGAGATGAAC AAGTCGAATA
(RS.A2)   AACCTCGCCA AACATCAACG GAGTTTGGGT TATGATGGAT GGAGATGAAC AAGTCGAATA
(AL.23C)  AACCTCGCCA AACATCAACG GAGTTTGGGT TATGATGGAT GGAGATGAAC AAGTCGAATA
(TQ.X2C)                                                          CGAATA
(TQ.X6)                                                           CGAATA
(CON)     AACCTCGCCA AACATCAACG GAGTTTGGGT TATGATGGAT GGAGATGAAC AAGTCGAATA

                430         440         450         460         470         480
(TQ.X2C)  CCCACTAAAA CCAATC4T2G A43A2GCAAA ACCAACA1TT AGGCAAATCT AGGCACATTT
(TQ.X6)   CCCACTAAAA CCAATCGTTG AGAATGCAAA ACCAACA1TT AGGCAAATCA TGGCACATTT
(AL.X14)  CCCACTAAAA CCAATCGTTG AGA
(AL.X2)   CCCACTAAAA CCAATCGTTG AGAATGCAAA ACCAACACTT AGGCAAATCA TGGCACATTT
(RS.A2)   CCCACTAAAA CCAATCGTTG AGAATGCAAA ACCAACACTT AGGCAAATCA TGGCACATTT
(AL.23C)  CCCACTAAAA CCAATCGTTG AGAATGCAAA ACCAACACTT AGGCAAATCA TGGCACATTT
(CON)     CCCACTAAAA CCAATCGTTG AGAATGCAAA ACCAACACTT AGGCAAATCA TGGCACATTT

                490         500         510         520         530         540
(TQ.X2C)  CTCAGATGTT GCAGAAGCGT ATATAGAAAT GCGCAACAAA AAGGAACCAT ATATGCCACG
(TQ.X6)   CTCAGATGTT GCAGAAGCGT ATATAGAAAT GCGCAACAAA AAGGAACCAT ATATGCCACG
(HP.13C)  CTCAGATGTT GCAGAAGCGT ATATAGAAAT GCGCAACAAA AAGGAACCAT ATATGCCACG
(AL.X2)   CTCAGATGTT GCAGAAGCGT ATATAG333T GCGCAACA33 33GG33CCAT ATATGCCACG
(RS.A2)   CTCAGATGTT GCAGAAGCGT ATATAGAAAT GCGCAACAAA AAGGAACCAT ATATGCCACG
(AL.23C)  CTCAGATGTT GCAGAAGCGT ATATAGAAAT GCGCAACAAA AAGGAACCAT ATATGCCACG
(CON)     CTCAGATGTT GCAGAAGCGT ATATAGAAAT GCGCAACAAA AAGGAACCAT ATATGCCACG

                550         560         570         580         590         600
(TQ.X2C)  ATATGGTTTA GTTCG2AATC TGCGCGTAGG AAGTTTGGCT CGCTATGCTT TTGACTTTTA
(TQ.X6)   ATATGGTTTA GTTCGTAATC TGCGCGATGG AAGTTTGGCT CGCTATGCTT TTGACTTTTA
(HP.13C)  ATATGGTTTA GTTCGTAATC TGCGCGATGG AAGTTTGGCT CGCTATGCTT TTGACTTTTA
(AL.X2)   ATATGGTTTA GTTCG3T3CT GTCGCGATGG 3AGTTTGGCT 1GC2ATGCTT TTG3CTTTTT
(RS.A2)   ATATGGTTTA GTTCGTAATC TGCGCGATGG AAGTTTGGCT CGCTATGCTT TTGACTTTTA
(AL.23C)  ATATGGTTTA GTTCGTAATC TGCGCGATGG AAGTTTGGCT CGCTATGCTT TTGACTTTTA
(CON)     ATATGGTTTA GTTCGTAATC TGCGCGATGG AAGTTTGGCT CGCTATGCTT TTGACTTTTA

                610         620         630         640         650         660
(AL.X2)   GAXGTTCACT CXXGXGACAC XXGTXXGGGC TAGXGXGGCC AAXATTCAAA TG3AGGCC
(TQ.X2C)  TGAAGTTACA TCACGGACAC CAGTGAGGGC TAGAGAGGCA CACATTCAAA TGAAGGCCGC
(TQ.X6)   TGAAGTTACA TCACGGACAC CAGTGAGGGC TAGAGAGGCA CACATTCAAA TGAAGGCCGC
(HP.13C)  TGAAGTTACA TCACGGACAC CAGTGAGGGC TAGAGAGGCA CACATTCAAA TGAAGGCCGC
(HA.B1)                                                                 GC
(RS.A2)   TGAAGTTACA TCACGGACAC CAGTGAGGGC TAGAGAGGCA CACATTCAAA TGAAGGCCG1
(AL.23C)  TGAAGTTACA TCACGGACAC CAGTGAGGGC TAGAGAGGC
(CON)     TGAAGTTACA TCACGGACAC CAGTGAGGGC TAGAGAGGCA CACATTCAAA TGAAGGCCGC
```

```
                670         680         690         700         710         720
(AL.X1C)                 CTCAAT CTCGACTTTT CGGATTGGAT GGTGGCATTA GTACACAAGA
(AL.X6R)    C2TTAAAA TCAGCTCAAT CTCGACTTTT CGGATTGGAT GGTGGCATTA GTACACAAGA
(HP.13C)  AGCTT2AAAA TCAGCTCAAT CTCGACTTTT CGGATTGGAT GGTGGCATTA GTACACAAGA
(AL.X3A)                 CTCAAT CTCGACTTTT CGGATTGGAT GGTGGCATTA GTACACAAGA
(RS.C2C)                                                          ACACAAGA
(RS.C3)                                                           ACACAAGA
(HA.B1)   AG1TTTAAAA T1AGCTCAAT CTCGACTTTT 1GGATTGGAT GGTGGCATTA GTA1A1AAGA
(AL.X3)                  CTCAAT CTCGACTTTT CGGATTGGAT GGTGGCATTA GTACACAAGA
(RS.A2)   AGCTTTAAAA TCAGCTCAAT CTCGACTTT
(TQ.X2C)  AGCTTTAAAA TCAGCTCAAT CTC
(TQ.X6)   AGCTTTAAAA TCA4CTCAAT CTCGA
(CON)     AGCTTTAAAA TCAGCTCAAT CTCGACTTTT CGGATTGGAT GGTGGCATTA GTACACAAGA

                730         740         750         760         770         780
(AL.X1C)  GGAAAACACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG
(AL.X6R)  GGAAAACACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG
(HP.13C)  GGAAAACACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG
(AL.X3A)  GGAAAACACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG
(HA.B1)   GGAAAAXXXA GAGAGGCA1A CCA1CGAGGA TGTTTXTCCA AGTATG1ATA 1T1TA1TTGG
(AL.X3)   GGAAAACACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG
(RS.C3)   GGAAAACACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG
(RS.C2C)  GGAAAXCACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG
(CON)     GGAAAACACA GAGAGGCACA CCACCGAGGA TGTTTCTCCA AGTATGCATA CTCTACTTGG

                790         800         810         820         830         840
(AL.X1C)  AGTGAAGAAC ATGTGATTGT AGTGTCTTTC CGGACGATAT ATAGATATTT ATG22TGCAG
(AL.X6R)  AGTGAAGAAC ATGTGATTGT AGTGTCTTTC CGGACGATAT ATAGATATTT ATG22TGCAG
(HP.13C)  AGTGAAGAAC ATGTGA
(AL.X3A)  AGTGAAGAAC ATGTGATTGT AGTGTCTTTC CGGACGATAT ATAGATATTT ATGTTTGCAG
(HA.B1)   AGTGAAGAA1 AT4TGATTGT AGTGT1TTT1 1GGACGATAT ATAGATATTT ATGTTTG1AG
(RS.C3)   AGTGAAGAAC ATGTGATTGT AGTGTCTTTC CGGACGATAT ATAGATATTT ATGTTTGCAG
(RS.C2C)  AGTGAAGAAC ATGTGATTGT AGTGTCTTTC CGGACGATAT ATAGATATTT ATGTTTGCAG
(AL.X3)   AGTGAAGAAC ATGTGATTGT AGTGTCTTTC CGGACGATAT ATAGATATTT ATGTTTGCAG
(HP.C19)                                  GGACGATAT ATAGATATTT ATGTTTGCAG
(HP.X9C)                                  GGACGATAT ATAGATATTT ATGTTTGCAG
(CON)     AGTGAAGAAC ATGTGATTGT AGTGTCTTTC CGGACGATAT ATAGATATTT ATGTTTGCAG

                850         860         870         880         890         900
(HP.X9C)  TAAGTATTTT GGCTTTTCCT GTACTACTTT TATCGAAATT AATAATCGTT 2GAATATTAC
(HP.C19)  TAAGTATTTT GGCTTTTCCT GTACTACTTT TATCGAAATT AATAATCGTT TGAATATTAC
(AL.X1C)  TAAGTATTTT GGCTTTTCCT GTACTACTTT TATCGAAATT AATAATCGTT TGAATATTAC
(AL.X6R)  TAAGTATTTT GGCTTTTCCT GTACTACTTT TATCGAAATT AATAATCGTT TGAATATTAC
(TQ.C22)                                  ATCGAAATT AATAATCGTT TGAATATTAC
(AL.X3A)  TAAGTATTTT GGCTTTTCCT GTACTACTTT TATCGAAATT AATAATCGTT TGAATATTAC
(HA.B1)   TAAGTATTTT GGCTTTTCCT GT
(RS.C3)   TAAGTATTTT GGCTTTTCCT GT
(RS.C2C)  TAAGTATTTT GGCTTTTCCT GT
(AL.X3)   TAAGTATTTT GGCTTTTCCT GTACTACTTT TATCGAAATT AATAATCGTT TGAATATTAC
(RS.A6)                      ACTACTTT TATCGAAATT AATAATCGTT TGAATATTAC
(TQ.X3C)                                TCGAAATT AATAATCGTT TGAATATTAC
(RS.B3)                      ACTAC222 2ATCGAAA22 AA2AA2CGTT TGAATATTAC
(CON)     TAAGTATTTT GGCTTTTCCT GTACTACTTT TATCGAAATT AATAATCGTT TGAATATTAC
```

```
                910         920         930         940         950         960
(TQ.X3C)  TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAGCTG TCGTTTCTGT
(HP.X9C)  TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAGCTG 2CGTTTCTGT
(HP.C19)  TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAGCTG 2CGTTTCTGT
(AL.X1C)  TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAG
(AL.X6R)  TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAG
(TQ.C22)  TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAGCTG TCGTTTCTGT
(AL.X3A)  TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAG
(AL.X3)   TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAG
(XB.18C)                                  CGTTGTAXTG ACCTTAGCTG XCGXXXCTGT
(RS.A6)   TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAGCTG TCGTTTCTGT
(RS.B3)   TGGCAGATAG GGGTGGTATA GCGATTCCG2 CGTTGTAGTG ACCTTAGCTG TCG222CTGT
(CON)     TGGCAGATAG GGGTGGTATA GCGATTCCGT CGTTGTAGTG ACCTTAGCTG TCGTTTCTGT

                970         980         990        1000        1010        1020
(RS.A6)   ATTATTATGT TTGTATAAAA GTGCCGGGTT GTTGTTGTTG TGGCTGATCT ATCGATTAGG
(XB.18C)  ATTATTATGT TTGTACAAAA 1TGCCGGGTT GTTGTTGTTG TGGCTGATCT ATCGATTAXX
(AL.18)          T  TTGTATAAAA GTGCCGGGTT GTTGTTGTTG TGGCTGATCT ATCGATTAGT
(RS.B3)   ATTATTATGT TTGTATAAAA GTGCCGGGTT GTTGTTGTTG TGGCTGATCT ATCGATTAGT
(AL.13B)     ATYAT4X XTGTATAAAA GTGCCGVXCT 4XT4XT41T4 TVXCTGATCT ATCGA22AG2
(TQ.X3C)  ATTATTATGT TTGTATAAAA GTGCCGGGTT GTTGTTGTTG TGGCTGATCT ATCGA
(HP.X9C)  ATTATTATGT TTGTATAAAA GTGCCGGG
(HP.C19)  ATTATTATGT TTGTATAAAA GTGCC
(TQ.C22)  ATTATTATGT TTGTATAAAA GTGCCGGGTT GT
(CON)     ATTATTATGT TTGTATAAAA GTGCCGGGTT GTTGTTGTTG TGGCTGATCT ATCGATTAGT

               1030        1040        1050        1060        1070        1080
(XB.18C)  TGATGXTGCG ATTXGTXGTA GCAGTGACTA TGTCTGGATT TAGTTACTTG GGTGATGCTG
(RS.A6)   TGATGTTGCG ATTTGTCG2A GCAG2GACTA TGACTGGA22 2AG22ACTTG GGXGATGCTG
(AL.18)   T8ATGTTGCG ATTTGTCGTA GCAGTGACTA TGTCTGGATT TAGTTACTTG GGTGATGCTG
(RS.B3)   TGATGTTGCG ATTTGTCGTA GCAGTGACTA TGTCTGGATT TAGTTACTTG GGTGATGCTG
(AL.13B)  TGATG22GCG A22T4TCGTA GCA4TGACT3 TG2CTGGATT TAGTTACTTG GGTGTAGCTG
(TQ.X14)                TTGTCGTA GCAGTGACTA TGTCTGGATT TAGTTACTTG GGTGATGCTG
(CON)     TGATGTTGCG ATTTGTCGTA GCAGTGACTA TGTCTGGATT TAGTTACTTG GGTGATGCTG

               1090        1100        1110        1120        1130
(TQ.X14)  TGATTCTGTC ATAGCAGTGA CTGTAAACTT CAATCAGGAG ACAAAAAAAA AAAA
(XB.18C)  TGATTCTGTC ATAGCAGTGA CTGTAAACTT CAATCAGGAG ACAAAAAAAA AAAA
(RS.A6)   TGATXCTGTG A2AGCAG2GA CTGTAAACTT CAATCAGGAG ACAAAAAAAA AAAA
(AL.18)   TGATTCTGTC ATAGCAGTGA CTGTAAACTT CAATCAGGAG ACA
(RS.B3)   TGATTCTGTC ATAGCAGTGA CTGTAAACTT CAATCAGGAG ACAAAAAAAA AA
(AL.13B)  TGA221TGTC ATAGCAGTGA CTGTAAACTT CAATCAGGAG ACAAAAAAAA AAAA
(CON)     TGATTCTGTC ATAGCAGTGA CTGTAAACTT CAATCAGGAG ACAAAAAAAA AAAA
```

APPENDIX C: RESTRICTION ENDONUCLEASE SITES IN PVY^N COAT PROTEIN GENE.

Restriction endonuclease sites and fragment lengths resulting from the digestion of the PVY^N coat protein gene are listed. Restriction endonucleases with sites >6 base pairs are also included. The coat protein gene nucleotide sequence has the appropriate four and six base pair restriction sites written below.

Number	Enzyme	Site	Position	Length
1	CLAI	ATCGAT	7	6
2	TAQI	TCGA	7	0
3	SACIII	ACGT	92	85
4	ALUI	AGCT	138	46
5	SACIII	ACGT	145	7
6	AVAI	CctCGagG	204	59
7	XHOI	CTCGAG	204	0
8	TAQI	TCGA	205	1
9	AVAI	CctCGagG	241	36
10	XHOI	CTCGAG	241	0
11	TAQI	TCGA	242	1
12	SACIII	ACGT	268	26
13	RSAI	GTAC	283	15
14	TAQI	TCGA	415	132
15	DDEI	CTnAG	459	44
16	DDEI	CTnAG	482	23
17	MSTI	TGCGCA	512	30
18	HHAI	GCGC	514	2
19	HHAI	GCGC	565	51
20	THAI	CGCG	565	0
21	HAEIII	GGCC	657	92
22	FNU4HI	GCnGC	658	1
23	FNU4HI	GCnGC	661	3
24	ALUI	AGCT	663	2
25	AHAIII	TTTAAA	667	4
26	ALUI	AGCT	675	8
27	TAQI	TCGA	683	8
28	RSAI	GTAC	713	30
29	AVAIII	ATGCAT	767	54
30	HPAII	CCGG	811	44
31	RSAI	GTAC	863	52
32	TAQI	TCGA	874	11
33	SSPI	AATATT	896	22
34	HINFI	GAnTC	924	28
35	DDEI	CTnAG	944	20
36	ALUI	AGCT	948	4
37	HPAII	CCGG	985	37
38	SCRFI	CCnGG	986	1
39	SAU3A	GATC	1006	20
40	CLAI	ATCGAT	1013	7
41	TAQI	TCGA	1013	0
42	HINFI	GAnTC	1083	70

Number	Enzyme	Site	Position	Length
1	FOKI	GGATGnnnnnnnnn	45	44
2	BSMI	GAATGCn	163	118
3	MNLI	CCTCnnnnnnn	374	211
4	TTH111II	CAAagCAnnnnnnnnnnn	386	12
5	FOKI	GGATGnnnnnnnnn	411	25
6	BSMI	GAATGCn	449	38
7	PFLMI	CCAnnnnnTGG	543	94
8	BBVI	GCAGCnnnnnnnn	672	129
9	FOKI	GGATGnnnnnnnnn	711	39
10	TAQII	cgACCcgAnnnnnnnnnnn	759	48
11	FOKI	GGATGnnnnnnnnn	762	3
12	MBOII	GAAGAnnnnnnnn	797	35
13	TTH111I	GACnnnGTC	1050	253
14	HPHI	GGTGAnnnnnnnn	1084	34

```
       10         20         30         40         50         60         70
CACAATCGATGCAGGAGGAAGCACTAAAAAGGATGCAAAACAAGAGCAAGGTAGCATTCAACCAAATTTC
    CLAI
     TAQI

       80         90        100        110        120        130        140
AACAAGGAAAAGGAAAAGGACGTGAATGTTGGAACATCTGGAACTCATACTGTGCCACGAATTAAAGCTA
                  SACIII                                              ALUI

      150        160        170        180        190        200        210
ICACGTCCAAAATGAGAATGCCCAAGAGTAAAGGTGCAATTGCATTAAATTTGGAACACTTACTCGAGTA
  SACIII                                                          AVAI
                                                                   TAQI
                                                                  XHOI

      220        230        240        250        260        270        280
TGCTCCACAGCAAATTGACATCTCAAATACTCGAGCAACTCAATCACAGTTTGATACGTGGTATGAAGCA
                              AVAI                        SACIII
                               TAQI
                              XHOI

      290        300        310        320        330        340        350
GTACAACTTGCATACGACATAGGAGAAACTGAAATGCCAACTGTGATGAATGGGCTTATGGTTTGGTGCA
RSAI

      360        370        380        390        400        410        420
TTGAAAATGGAACCTCGCCAAACATCAACGGAGTTTGGGTTATGATGGATGGAGATGAACAAGTCGAATA
                                                                   TAQI

      430        440        450        460        470        480        490
CCCACTAAAACCAATCGTTGAGAATGCAAAACCAACACTTAGGCAAATCATGGCACATTTCTCAGATGTT
                                        DDEI                    DDEI

      500        510        520        530        540        550        560
GCAGAAGCGTATATAGAAATGCGCAACAAAAAGGAACCATATATGCCACGATATGGTTTAGTTCGTAATC
                  HHAI
                 MSTI

      570        580        590        600        610        620        630
TGCGCGATGGAAGTTTGGCTCGCTATGCTTTTGACTTTTATGAAGTTACATCACGGACACCAGTGAGGGC
 HHAI
  THAI

      640        650        660        670        680        690        700
TAGAGAGGCACACATTCAAATGAAGGCCGCAGCTTTAAAATCAGCTCAATCTCGACTTTTCGGATTGGAT
                        HAEIIIALUI          ALUI      TAQI
                         FNU4HI  AHAIII
                            FNU4HI

      710        720        730        740        750        760        770
GGTGGCATTAGTACACAAGAGGAAAACACAGAGAGGCACACCACCGAGGATGTTTCTCCAAGTATGCATA
         RSAI                                                      AVAIII

      780        790        800        810        820        830        840
CTCTACTTGGAGTGAAGAACATGTGATTGTAGTGTCTTTCCGGACGATATATAGATATTTATGTTTGCAG
                                        HPAII

      850        860        870        880        890        900        910
TAAGTATTTTGGCTTTTCCTGTACTACTTTTATCGAAATTAATAATCGTTTGAATATTACTGGCAGATAG
                  RSAI         TAQI                SSPI

      920        930        940        950        960        970        980
GGGTGGTATAGCGATTCCGTCGTTGTAGTGACCTTAGCTGTCGTTTCTGTATTATTATGTTTGTATAAAA
            HINFI                   ALUI
                                   DDEI

      990       1000       1010       1020       1030       1040       1050
GTGCCGGGTTGTTGTTGTTGTGGCTGATCTATCGATTAGTTGATGTTGCGATTTGTCGTAGCAGTGACTA
   HPAII                  SAU3ACLAI
   SCRFI                      TAQI

     1060       1070       1080       1090       1100       1110       1120
TGTCTGGATTTAGTTAGTTGGGTGATGCTGTGATTCTGTCATAGCAGTGACTGTAAACTTCAATCAGGAG
                                   HINFI

     1130
ACAAAAAAAAAAAA
```

APPENDIX D: ENZYMES AND BIOLOGICAL COMPOUNDS

PVY ELISA kit, Boehringer Mannheim

Oligo $(dT)_{12\text{-}18}$, Pharmacia, 1 mg ml^{-1}

M-MLV reverse transcriptase, Bethesda Research Laboratories, 200 units ml^{-1}

[methyl-^{3}H]dTTP, Amersham, 47 Ci $mmol^{-1}$

[α-^{35}S]dCTP, Amersham, 1000 Ci $mmol^{-1}$

[α-^{35}S]dATP, Amersham, 400 Ci ml^{-1}

dATP, dGTP, dCTP, dTTP, Boehringer Mannheim

ddGTP, ddATP, ddCTP, ddTTP, Boehringer Mannheim, 1 μmol

Synthetic *Xba*I linkers, New England Biolabs, 1 A_{260} unit

E. coli DNA polymerase I, New England Biolabs, 10 units μl^{-1}

E. coli RNase H, Amersham, 4 units μl^{-1}

BSA, Sigma Chemical Co., 10 mg ml^{-1} in dH_2O, acetylated

T4 DNA polymerase, Bethesda Research Laboratories, 5 units μl^{-1}

CIP (special molecular biology grade), Boehringer Mannheim, 28 units μl^{-1}

T4 polynucleotide kinase, Bethesda Research Laboratories, 1 unit μl^{-1}

T4 DNA ligase, Boehringer Mannheim, 1 unit μl^{-1}

DNase I (RNase-free), Bethesda Research Laboratories, 10 mg ml^{-1}

RNase A, Sigma Chemical Co., 10 mg ml^{-1} in dH_2O

RNase H, Amersham, 2000 units ml^{-1}

*BamH*I restriction endonuclease, Bethesda Research Laboratories, 10 units μl^{-1}

*Sca*I restriction endonuclease, Boehringer Mannheim, 12 units μl^{-1}

*Acc*I restriction endonuclease, Bethesda Research Laboratories, 5 units μl^{-1}

*Cla*I restriction endonuclease, Bethesda Research Laboratories, 4 units μl^{-1}

*Ssp*I restriction endonuclease, New England Biolabs, 6 units μl^{-1}

Klenow DNA polymerase I, Boehringer Mannheim, 1000 units ml^{-1}

M13 17mer sequencing primer, New England Biolabs, 500 pmol ml^{-1}

Anti-[sheep/goat] immunoglobulin alkaline phosphatase-conjugated, Silenus, 5 mg ml^{-1}

Proteinase K (lyophilised), Boehringer Mannheim

Cefotaxime, Roussel

Carbenicillin, di-sodium salt, Sigma Chemical Co.

Ampicillin, Sigma Chemical Co.

Kanamycin monosulfate, Sigma Chemical Co.

Chloramphenicol, Sigma Chemical Co.

DNA herring sperm (lyophilised), Boehringer Mannheim